AD_____

Award Number: W81XWH-06-1-0715

TITLE: Novel Strategies for the Treatment of Estrogen Receptor-Negative Breast Cancer

PRINCIPAL INVESTIGATOR: Corey Speers

CONTRACTING ORGANIZATION: Baylor College of Medicine
Houston, TX 77030

REPORT DATE: April 2009

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1 April 2009	Annual Summary	4 Sep 2006 – 3 Mar 2009

4. TITLE AND SUBTITLE

Novel Strategies for the Treatment of Estrogen Receptor-Negative Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-06-1-0715

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Corey Speers

E-Mail: cs138697@bcm.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Baylor College of Medicine
Houston, TX 77030

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Previous gene expression profiling studies of breast cancer have focused on the entire genome to identify genes differentially expressed between estrogen receptor (ER)-positive and ER-negative cancers. Here we report a distinct kinase gene expression profile that identifies ER-negative breast tumors and subsets ER-negative breast tumors into 4 distinct subtypes. Furthermore, we show that this specific kinase profile is validated in breast cancer cell lines and independent sets of human tumors. Kinase expression knock-down studies show that many of these kinases are essential for the growth of ER-negative, but not ER-positive, breast cancer cell lines. Finally, survival analysis of patients with breast cancer shows that the S6 kinase pathway signature of ER-negative cancer confers an extremely poor prognosis, while patients whose tumors express high levels of immunomodulatory kinases have a significantly better prognosis. This study identifies a list of kinases that are prognostic and may serve as druggable targets for the treatment of ER-negative breast cancer.

15. SUBJECT TERMS
estrogen receptor-negative breast cancer, genomic profiling, kinases

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	300	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

Although ER-positive breast cancers account for 60-70% of breast cancers, 30% of breast cancers are ER-negative and poorly responsive to traditional therapies *(1).* Selective estrogen receptor modulators (SERMs) (such as tamoxifen and raloxifene) and aromatase inhibitors reduce ER-positive breast cancer recurrence by approximately 50% (*2,3*). These agents, however, are not effective in treating or preventing ER-negative breast cancer. Currently, chemotherapy is used to treat ER-negative tumors *(4).* Such therapy is generally toxic and is not specifically targeted to ER-negative breast cancer, but instead non-specifically kills rapidly dividing cells. The only targeted therapy shown to be effective for a subset of ER-negative breast cancer is herceptin, a monoclonal antibody that only targets those tumors that overexpress the Her2 receptor *(4).* This information taken collectively demonstrates that to make additional advances in preventing and treating breast cancer, effective agents for ER-negative breast cancer must be developed.

It is evident that multiple signal transduction pathways play crucial roles in breast cancer development. The growth signal sensed by the cell is conveyed to the nucleus through interactions of proteins in series, each one activating another, through signal transduction pathways. Once the signal is received in the nucleus, transcription factors activate genes important for cell growth and survival. As was noted earlier, many of these pathways are understood in ER-positive cancers and have been the targets of small molecule inhibitors that can interrupt this mitogenic signaling, preventing and treating these cancers. Currently the mechanisms governing ER-negative breast cancer cell growth are unknown. It is clear that estrogen signaling is not the pathway that governs the mitogenic pathway, but despite the best efforts of numerous groups, the identification of pathways critical for ER-negative growth remains elusive. Recent advances in molecular biology have allowed for breakthroughs in the search for these growth pathways. Genome-wide expression arrays have allowed researchers to probe expression profiles in all different tissue types, including normal and malignant tissue *(5,6,7).* These studies, along with subsequent validation of their results, have led to advances in understanding breast cancer and have led to better tools for the clinician in evaluating patients with breast cancer. It is now possible to profile a tumor molecularly and determine what types of therapies will be most effective *(8,9,10).* Despite this increasing knowledge, however, it is clear that much work remains to be done.

BODY

1) Research Training Environment

The Breast Center at Baylor College of Medicine (BCM) provides a unique training environment with multiple opportunities for me to grow as a young research scientist. In the past year, I have taken full advantage of these opportunities as outlined:

- completed and received an "A" letter grade in the Molecular Carcinogenesis course taught here at BCM
- completed a course in Translational Breast Cancer Research, which is taught by faculty members of the Breast Center
- received permission to write my thesis
- presented data in poster format at the 2008 Annual Meeting of the American Association for Carncer Research, San Diego, CA, the graduate student symposium at BCM, the 6th Annual Dan L. Duncan Cancer Center, and gave oral presentations at the Medical Scientist Training Program Annual Retreat at Hotel Galvez, the 4th Annual Lester and Sue Smith Breast Center Retreat at Artesian Lakes, and the 2008 DOD Era of Hope Meeting in Baltmiore, MD.
- Received numerous first place awards for both oral and poster presentations at the aforementioned meetings.
- Co-Investigator in a currently accruing Phase II clinical trial entitled, "A Biologic Correlative Study of Dasatinib, a Multi-Targeted Tyrosine Kinase, in "Triple-Negative" Breast Cancer Patients"
- Attended and was awarded a travel grant to attend the Cold Spring Harbor Course entitled, "Integrated Data Analysis for High Throughput Biology"
- Received higher order bioinformatic training the in the labof Dr. John Quackenbush at Dana Farber Cancer Institute in Boston, MA

2) Research Project

Specific Aim 1: Identify novel targets for the treatment of ER-negative breast cancer using genomic analysis:

1.1) Identify the kinases and phosphatases that are differentially expressed in human ER-negative vs. ER-positive breast tumors using RNA affymetrix microarray chips.

1.2) Validate that the genes are differentially expressed in a second set of human ER-negative breast cancers using quantitative RT-PCR analysis.

1.3) For selected identified kinases or phosphatases determine whether these proteins are differentially expressed in ER-negative vs. ER-positive tumors

This aim has been completed and the results are reported in the attached paper, currently in review, entitled "Identification of Novel Kinase Targets for the Treatment of Estrogen Receptor-Negative Breast Cancer" (see **attached paper in appendix**)

Specific Aim 2: Identify novel targets for the treatment of ER-negative breast cancer using proteomic analysis:

2.1) Make protein lysates from the 110 human tumor samples (both ER-positive and ER-negative) and use these lysates in a reverse phase tissue lysate array. This is a quantitative automated proteomics assaying system that determines the expression level and activation status of signaling proteins. Using this array technology I will identify those signaling molecules that are differentially expressed between ER-negative and ER-positive tumors. Furthermore, I will assay for activation status of the molecule utilizing phospho-specific antibodies.

2.2) Validate that selected identified proteins or phosphoproteins are differentially expressed in a second set of human ER-negative breast tumors using western blotting.

Proteomic analysis identifies proteins and phospho-proteins that are differentially expressed between ER-positive and ER-negative disease. We are in the final stages of generating figures and will submitting a manuscript describing the results of these studies. Our studies do indicate that there is a high degree of correlation between protein and RNA expression (see **attached paper in appendix)**.

Specific Aim 3: Determine whether inhibition of the identified RNA and protein targets suppresses ER-negative breast cancer growth *in vitro* and *in vivo*.

3.1) For *in vitro* studies I will determine whether inhibition of signaling molecule function using siRNA knockdown inhibits ER-negative breast cancer cell growth. For these experiments I will use ER-negative cell lines selected to accurately represent *in vivo* breast cancers. These cells will be transfected with siRNA designed against signaling molecules identified in **Aim 1** and **2.** I will then use MTT, soft agar growth, and invasion assays to determine whether specific gene knockdown inhibits growth or invasion.

3.2) Use existing small molecule inhibitors of the identified signaling molecules to block the activity of these proteins and assay for growth suppression.

3.3) For *in vivo* studies, I will determine whether stably transfected shRNA or small molecule inhibitors can suppress the growth of breast tumors when xenografted into nude mice.

To evaluate the role of kinase function in both ER-negative and ER-positive breast cancer growth, cell lines representing these two types of cancers (MDA-MB-468 and MDA-MB-231; and MCF-7 and T47D cells, respectively) were used in the siRNA experiments. Several of the kinases evaluated to date have significant growth inhibitory phenotypes when knocked down in ER-negative breast cancer cell lines, while they show no growth inhibitory effects in ER-positive cell lines (**see attached submitted paper in the appendix**). Also, progress is currently being made in evaluating the effect of all 37 kinases indentified in the intersection in Aim 1 in further siRNA knockdown experiments. Additionally, we are beginning to plan *in vivo* mouse xenograft experiments that will be conducted in the next 6 months. IRB-approval for such experiments has already been obtained and the mice are currently being obtained. Finally, based on the results of these studies, we are proceeding with a phase II clinical trial is IRB approved and now accruing patients to evaluate whether the multi-kinase inhibitor dasatinib can regress advanced triple negative breast tumors in stage II and III women with triple negative disease.

KEY RESEARCH ACCOMPLISHMENTS

- Gene expression microarray analysis is a robust means of identifying kinases upregulated in ER-negative breast tumors.
- Unsupervised clustering analysis identifies 4 distinct subsets of ER-negative breast cancer

- Identified kinases can be validated using Q-PCR and western blot analysis (data not shown) in both breast cancer cell lines and human breast tumors
- Tumors from these different kinase clusters have different metastasis free and overall survival and may identify patients that require more aggressive adjuvant therapy
- *In silico* promoter analysis identifies E2F1 as a regulator of expression in one of the 4 identified subsets of ER-negative breast cancer
- Inhibition of the several identified kinases using siRNA inhibits ER-negative, but not ER-positive, breast cancer cell growth *in vitro*
- Based on the kinase targets identified in this research, the multi-kinase inhibitor dasatinib, which targets many of the kinases identified in this analysis, is being taken into a phase II clinical trial that we are heading.
- RPPA identifies proteins and pathways upregulated in ER-negative breast cancer which can be validated in independent datasets
- RPPA analysis identifies high expression of cyclin B1 and PAI1 protein as poor prognostic markers.
- Intrinsic subtypes of breast cancer can be identified by protein expression (not just RNA)
- RNA and protein expression correlate very well between genomic and proteomic arraying platforms

REPORTABLE OUTCOMES

- see attached publications currently in review (summarized in final thesis)

CONCLUSION

In this report we show that Affymetrix gene expression profiling of human breast tumors is able to identify kinases that are differentially-expressed in ER-negative breast cancers as compared to ER-positive breast cancers. Further analysis revealed that ER-negative tumors can be clustered into 4 distinct groups, depending on the specific kinases expressed and the level of their expression. Analysis of publicly available breast tumor data sets confirmed that these kinases are indeed upregulated in ER-negative breast cancer. Studies in which knock-down of selected kinases using siRNA were conducted and demonstrated that many of the identified kinases are critical for ER-negative, including "triple-negative", breast cancer growth. Finally, analysis of kinase expression in human breast tumors demonstrated that the individual subtypes of ER-negative breast cancer identified by their kinase profile here have different prognoses. Specifically, these studies demonstrate that ER-negative tumors that express highly the kinases from the S6 kinase group have a particularly bad prognosis, while those that express immunomodulatory kinases have a relatively good prognosis. Such results suggest that characterization of human tumors based on kinase expression can be used to select patients appropriate for novel therapies. In addition, this study identifies novel targets for the treatment of ER-negative breast cancer, including the aggressive "triple-negative" form of breast cancer.

REFERENCES:

(*1*) Society AC. Estimated new cancer cases and deaths by gender, US, 2002. Cancer facts & figures 2002 2002:4.

(*2*) Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. (1998). "Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study." J Natl Cancer Inst, 90:1371-88.

(*3*) Cummings SR, Eckert S, Krueger KA, Grady D, Powles TJ, Cauley JA, et al. (1999). "The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation." JAMA 281:2189-97.

(4) Tan, AR., S.M. Swain, et al. (2001). "Adjuvant chemotherapy for breast cancer: an update." Semin. Oncol. (4): 359-376.

(5) Zhao, H.,A. Langerod, et al. (2004). "Different gene expression patterns in invasive lobular and ductal carcinomas of the breast." Mol Biol Cell 15(6): 2523-36

(6) Wilson, C.A., and J. Dering (2004). "Recent translational research: microarray expression profiling of breast cancer—beyond classification and prognostic markers?" Breast Cancer Res 6(5): 192-200

(7) West, M., C. Blanchette, et al. (2001). "Predicting the clinical status of human breast cancer by using gene expression profiles." Proc Natl Acad Sci USA 98(20): 11462-7.

(8) Tan, Y., L. Shi, et al. (2005). "Multi-class cancer classification by total principal component regression (TPCR) using microarray gene expression data." Nucleic Acids Res 33(1): 56-65

(9) Subramaniam, D.S. and C. Isaacs (2005). "Utilizing prognostic and predictive factors in breast cancer." Curr Treat Options Oncol 6(2): 147-59.

(10) van't Veer, L.J., H. Dai, et al. (2002) "Gene expression profiling predicts outcomes of breast cancer." Nature 415(6871): 530-6.

Chapter 1

Introduction and Background

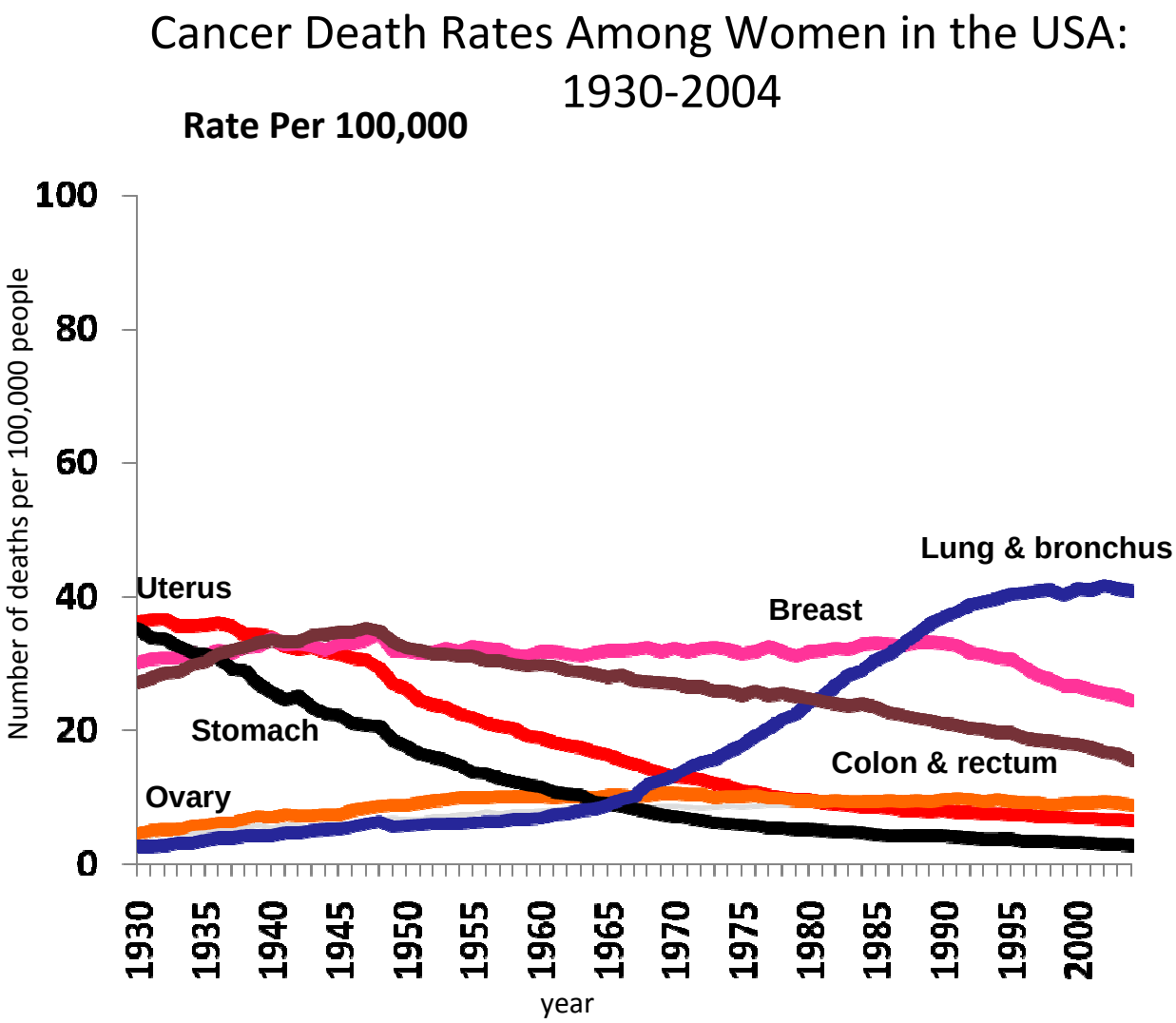
1.1 Breast Cancer Facts and Figures

Breast cancer is the second leading cause of cancer-related deaths among women in the United States. It is estimated that in 2008 alone, over 211,000 women will be diagnosed with breast cancer, and 40,000 women will die from its effects [1]. Despite President Nixon's declaration of war on cancer over 30 years ago, deaths related to breast cancer remained relatively unchanged until the early 1990s (**Figure 1.1**). Advances in newer and better methods of detection, as well as breakthroughs in prevention and targeted therapy have begun to decrease breast cancer mortality. Despite these modest decreases, however, it is clear that better and more effective breast cancer therapies need to be developed.

1.2 Development of Targeted Therapies for Breast Cancer

Although the majority of human breast tumors are estrogen receptor alpha (hereafter referred to as ER)-positive and respond well to selective estrogen receptor modulators (SERMs) or aromatase inhibitors (Als), approximately 30-40% of breast tumors are estrogen receptor alpha (hereafter referred to as ER)-negative. These ER-negative tumors are usually more aggressive and account for a much great proportion of deaths than their prevalence would suggest. This is primarily because of poor tumor response to traditional therapies [2]. Breast cancer development and progression involves that interplay between two major classes of growth-promoting agents, steroid hormones (estrogens and progestins) and polypeptide growth factors. These growth

Figure 1.1



Age-adjusted to the 2000 US standard population.
Source: US Mortality Data 1960-2004, US Mortality Volumes 1930-1959, National Center for Health Statistics, Centers for Disease Control and Prevention, 2008.

Figure 1.1 Age Adjust Cancer Rates: Graph shows age adjusted rates of cancer deaths by disease over time through 2005. Figure obtained with permission from the American Cancer Society. *Cancer Facts & Figures 2008.* Atlanta: American Cancer Society; 2008.

factors and their receptors have served as important targets for the treatment of ER-positive and HER2-positive breast cancer. Estrogen receptor refers to a group of receptors which are activated by the hormone 17β-estradiol (estrogen). There are two types of estrogen receptors, one encoded by the gene *ESR1* on chromosome six and commonly referred to as estrogen receptor-alpha (ERα), and the other, encoded by the gene *ESR2* on chromosome fourteen, commonly referred to as estrogen receptor-beta (ERβ). While these nuclear hormone receptors may exist as homodimers or heterodimers, the tissue expression of these different forms of estrogen receptor differ. ERα is found in mostly in the endometrium, breast, ovarian, and hypothalamic tissue [3]. The expression of the ERβ protein has been documented in kidney, brain, bone, heart, and lung tissue, as well as in the intestinal mucosa, prostate, and endothelial cells [3]. Because of the differing effects of these receptors and the predominant expression of ERα in the breast, the remainder of this thesis will concern itself only with this form of the estrogen receptor. Indeed, the importance of ERα was demonstrated by the development of anti-estrogens (anti ERα drugs) such as tamoxifen and fulvestrant, and more recently aromatase inhibitors, that have led to significant improvements in the overall survival of patients with ER-positive breast cancer [4-8]. Similarly targeting HER2/*neu* (ErbB2), which is overexpressed on 20-30% of breast cancers, with trastuzumab and more recently lapatinib has resulted in a marked improvement in both response and survival [9-15]. Unfortunately, only 30-60% of women with HER2-positive tumors benefit from trastuzumab [13-15]. Recent efforts utilizing an integrated

genomic and proteomic approach shows that aberrations in the phosphatidylinositol-3 kinase pathway identifies patients likely to have a limited response to trastuzumab [16]. Despite the success of anti-HER2 therapy, the lack of clear targets for ER-negative cancers, including those patients with HER2-positive tumors who fail therapy, is sorely needed to improve patient outcomes.

1.3 Development of Peptide Growth Factor Receptor Inhibitors for the Treatment of Breast Cancer

Peptide growth factors or their receptors are currently being investigated as possible targets for the treatment of breast cancer. These include the epidermal growth factor receptor (EGFR), human epidermal growth factor receptor type 2 (HER2), insulin-like growth factor receptor (IGFR), fibroblast growth factor receptors (FGFR) and vascular endothelial growth factor (VEGF). In preclinical studies these pathways have been shown to be active in breast cancer cells, and inhibitors of these pathways are being tested in both preclinical and clinical trials.

1.4 HER2 and EGFR Inhibition for the Treatment of Breast Cancer

The ErbB receptor family consists of four trans-membrane receptor tyrosine kinases (RTKs), two of which are frequently misregulated in breast cancer and associated with a poor prognosis: the epidermal growth factor receptor (EGFR or ErbB1) and the human epidermal growth factor receptor type 2 (HER2 or ErbB2) [17]. Ligand

binding to the extracellular domain of monomeric EGFR leads to homodimerization or heterodimerization, followed by autophosphorylation of the intracellular kinase domains [17]. HER2 does not have a known ligand; it is transactivated following dimerization and is the preferred heterodimer partner for EGFR [17]. Both EGFR and HER2 are, therefore, valid therapeutic targets in breast cancer research. In the past decade, a monoclonal antibody directed against HER2 has been developed and used clinically to treat patients whose tumors overexpress HER2. The clinical success of the monoclonal antibody trastuzumab provides strong evidence of the important role that HER2 has in the pathogenesis of breast cancer. Trastuzumab targets the extracellular domain of HER2 and has demonstrated efficacy in HER2-overexpressing metastatic and early stage breast cancers [18-22]. Although clearly a major advance in the treatment of HER2-positive breast cancer, as with most anticancer drugs, not all patients tolerate or respond to trastuzumab. Although trastuzumab significantly increased the response rate when combined with chemotherapy for the first-line treatment of metastatic breast cancer, some patients' tumors progressed, suggesting the existence of intrinsic resistance [18]. Moreover, the majority of patients who initially did respond developed disease progression within 1 year while still receiving trastuzumab, suggesting the emergence of secondary resistance. Furthermore, the use of trastuzumab has led to concerns regarding cardiotoxicities, especially when the treatment is combined the anthracyclines in the adjuvant chemotherapy setting [23, 24].

Because of inherent or acquired resistance to treatment with the monoclonal antibody trastuzumab, alternative strategies of targeting the EGFR and HER2 signaling pathways have been developed. EGFR inhibitors such as gefitinib and erlotinib, and the dual EGFR/HER2 inhibitor lapatinib have also shown promise in early clinical trials [4-7]. Lapatinib (Tykerb®) is an orally active, small molecule inhibitor of EGFR and HER2 tyrosine kinases that was approved in the US in 2007 for use in combination with capecitabine for the treatment of patients with advanced or metastatic HER2-positive breast cancer who have received prior therapy including an anthracycline, a taxane, and trastuzumab. Lapatinib interrupts signal transduction from the HER2 and EGFR receptors by competing with ATP for the intracellular ATP-binding domain of these RTKs [25]. This mechanism is distinct from that of trastuzumab which is a monoclonal antibody that targets the extracellular domain of the HER2 receptor. The intracellular mechanism of action of lapatinib provided the rationale for evaluating lapatinib in trastuzumab-resistant patients. Studies in which women with metastatic breast cancer were treated by lapatinib showed that the drug prolonged time to progression by 17.2 weeks (19.7 to 36.9 wks) without inducing the cardiotoxicity seen with trastuzumab [9]. These and other results led to FDA-approval of lapatinib for the treatment of HER2-positive metastatic breast cancer. However, not all HER2-positive patients respond to this therapy. To elucidate which tumor types respond to lapatinib effectively, neo-adjuvant trials are currently being conducted in women with stage III and IV breast cancer using treatment with lapatinib. Lapatinib is also being tested as adjuvant therapy

for the treatment of early stage HER2-positive breast cancer in the ongoing TEACH trial [26]. In addition, a trial using lapatinib in women with HER2-positive or EGFR-positive ductal carcinoma in situ (DCIS) is currently underway. These studies will elucidate the efficacy of lapatinib in treating specific stages of HER2- and EGFR-positive breast cancer. However, it is clear that some HER2-positive breast cancers are resistant to this therapy and that additional agents either as single agents or in combination with trastuzumab or lapatinib will be needed to treat these tumors.

1.5 VEGF Inhibitors for the Treatment of Breast Cancer

Angiogenesis (the formation of new blood vessels) is important in the growth and progression of solid tumors, including breast cancer. The main pro-angiogenic factor, vascular endothelial growth factor (VEGF), is a potent angiogenic cytokine that induces mitosis and also regulates the permeability of endothelial cells. An increase in VEGF expression in tumor tissue has been found in solid malignancies and is associated with metastasis formation and poor prognosis [27, 28]. Bevacizumab, a recombinant humanized monoclonal antibody developed against VEGF, binds to soluble VEGF, preventing receptor binding and inhibiting endothelial cell proliferation and vessel formation. Pre-clinical and clinical studies have shown that bevacizumab alone or in combination with a cytotoxic agent decreases tumor growth and increases median survival time and time to tumor progression [29, 30]. Combination therapy comprising bevacizumab with paclitaxel recently received accelerated approval from the US Food

and Drug Administration (FDA) for use in the first-line treatment of patients with metastatic breast cancer [31-33]. After proving to have tolerable toxicities in first-line treatment of patients with breast cancer in phase I and II trials, phase III trials were initiated. One such trial, E2100, provided the basis for FDA approval [32]. In this trial, bevacizumab (10 mg/kg on days 1 and 15) plus paclitaxel (90 mg/m2 days 1, 8, and 15 every 28 days) was given until disease progression and the combination therapy was shown to approximately double median progression-free survival (PFS; 11.8 months vs. 5.9 months; hazard ratio = 0.60; $P < .001$) compared with paclitaxel alone; by contrast, a statistically significant improvement in overall survival was not seen with the addition of bevacizumab, although post study analysis demonstrated a significant increase in 1-year survival for the combination arm [32]. Bevacuzimab is now also being tested in combination with erlotinib (an EGFR inhibitor) in a phase II trial of patients with metastatic breast cancer to see if anti-VEGF and anti-EGFR therapies are superior to monotherapy for the treatment of patients with metastatic breast cancer [34].

1.6 IGFR Inhibitors for the Treatment of Breast Cancer

The insulin-like growth factor pathway plays a major role in cancer cell proliferation, survival, and resistance to anti-cancer therapies in many human malignancies, including breast cancer. Several characteristics of the IGF signaling pathway make it an especially attractive target. The expression of IGF-1R, the major signal transducing receptor of the pathway, appears to be necessary for malignant

transformation in preclinical models [35]. Indeed, forced overexpression of IGF-1R increases the timing and frequency of tumor development in animal models [36]. Also, IGF-1-deficient mice have greatly reduced capacity to support tumor growth and metastasis [37]. Another important feature of the IGF system is its near ubiquitous presence in breast cancer, including expression of the IGF-1R [5]. Here, the expression of IGF-1R may approach 90% [38, 39]. Compared to ER-positive or HER2-positive breast cancers (which represent 60-70% and 15–25% of all breast cancers, respectively) this represents a much broader potential group of patients that may be candidates for targeted therapy.

There are three main ways in which IGF signaling has been interrupted. Strategies that inhibit ligand-receptor interactions have been developed. Indeed, receptor blockade with the use of monoclonal antibody therapies against the IGF-1R have been the most clinically investigated approach to date. Tyrosine kinase inhibition is another strategy being employed with several agents in clinical and preclinical development [40, 41]. Finally, ligand sequestration through the use of monoclonal antibodies against the ligand, IGF, is a third potential approach. As a key signaling component of IGF system, the IGF-1 receptor (IGF-1R) is the target of several investigational agents in clinical and pre-clinical development [36, 41, 42]. There is also abundant evidence of crosstalk between IGF-1R and ER [43]. Thus, combination therapy that targets both the IGF receptor the estrogen receptor (using tamoxifen or aromatase inhibitors) are currently being tested in clinical trials for the treatment of breast cancer.

Despite the potential utility of anti-IGF therapies, side effects of therapy remain a severe limitation. The IGF pathway, as alluded to above, is critical for normal metabolic and homeostatic processes. It is required for cell growth, development, and metabolism in normal as well as cancerous tissue, and this ubiquitous expression leads to severe adverse systemic side effects. Another major problem with these anti-IGF pathway therapies is they interfere with insulin signaling and cause hyper- or hypoglycemia and hyperinsulinemia. Though efforts are underway to develop drugs that circumvent this limitation, it is clear that other targeted treatment options need to be developed for the treatment of ER-negative breast cancer.

1.7 Additional Targets for the Treatment of ER-Negative Breast Cancer

There are currently several other molecular targets that are being investigated as appropriate targets for the treatment of breast cancer. In breast cancer, the phosphoinositide 3-kinase (PI3K) pathway represents one such target. Activation of the PI3K/Akt/mTOR pathway may occur through activation of membrane receptors, including growth factors and the estrogen receptor. This pathway has been linked to promotion of survival in breast cancer cells, and resistance to chemotherapy, trastuzumab and tamoxifen [44-46]. Approximately 50% of patients with breast cancer have a mutation or loss of at least one copy of the PTEN gene, resulting in activation of PI3K signaling [47]. Preclinical studies have shown that in breast cancer cells with reduced PTEN expression, the PI3K/Akt/mTOR pathway becomes a fundamental

pathway for tumor proliferation and survival [48-50]. These cells consequently display increased sensitivity to LY294002 and rapamycin (both PI3 kinase inhibitors) compared with PTEN-positive cells [50]. Inhibition of mTOR has also been shown to restore tamoxifen sensitivity in breast cancer cells with aberrant Akt activity, but only in ER-positive breast cancers [49].

Due to the extensive implication of PI3K/Akt/mTOR pathway aberrations in breast cancer, clinical trials have been developed and completed evaluating the effectiveness of pathway inhibitors (like rapamycin analogs) in treating breast cancer. In one multicenter phase II study, 106 women with advanced breast cancer refractory to anthracyclines and taxanes were treated with weekly i.v. rapamycin ester cell cycle inhibitor (CCI)-779 at doses of 75 or 250 mg. Response rates were seen in nine patients (8%), including one complete response (CR) and eight partial responses (PR). An additional 43 patients achieved stable disease (SD) for at least 8 weeks for a total clinical benefit of 49% [51]. Based on other preclinical findings suggesting an association between hormone resistance and activation of the PI3K/ Akt/mTOR pathway, a phase II trial evaluating the combination of CCI-779 and the aromatase inhibitor letrozole was initiated. This study showed the treatment was well tolerated, and results were initially promising enough to move into phase III trials [52]. Despite early promise, drug resistance to targeted therapy remains a significant challenge to PI3K/Akt/mTOR pathway inhibition. Compensatory mechanisms, including increased expression of Akt in response to pathway inhibition, remain a shortcoming of such therapy [16, 53].

Additionally, PI3K/ Akt/mTOR pathway inhibition seems to be most efficacious in ER-positive breast cancers [54, 55]. Thus, the need to develop specific, effective therapies for ER-negative breast cancers remains.

1.8 The Need for Additional Targeted Therapies in ER-negative Breast Cancer

ER-negative breast cancer, and especially "triple-negative" breast cancer, that is, cancers lacking the expression of ER, PR and HER2, remain the most challenging type of breast cancer to manage. Currently, chemotherapy is used to ER-negative tumors [56]. Such therapy is toxic and is often not targeted, instead only non-specifically killing rapidly dividing cells. Because ER-negative breast cancers do not express many of the molecules that are the targets of previously discussed therapies, there is a critical need to identify additional molecular targets that can be specifically and effectively inhibited for the treatment of ER-negative breast cancer. Expression microarrays, which can identify molecules that play a role in the development and progression of ER-negative breast cancers, have the potential to identify such targets.

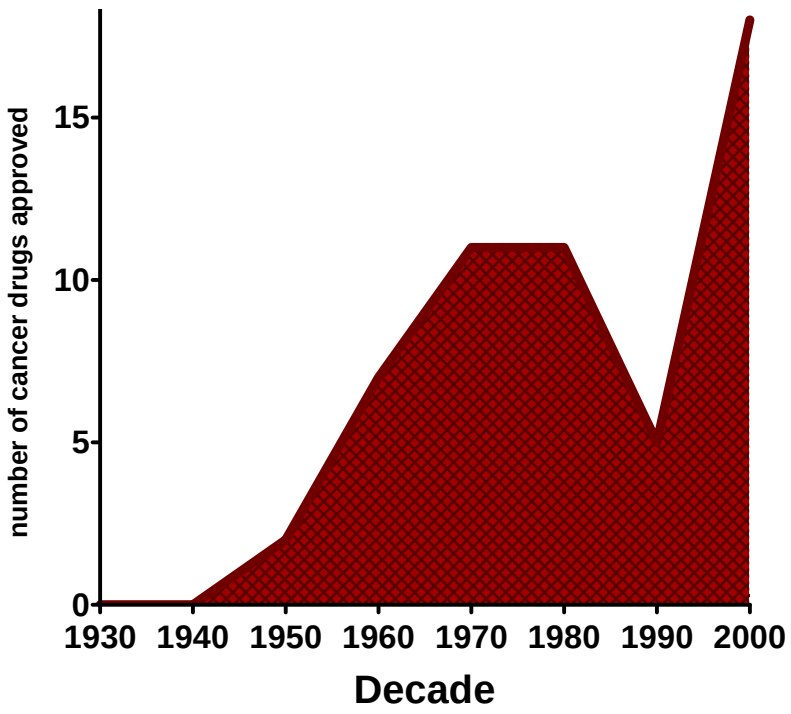
1.9 Gene Expression Profiling and the Development of Targeted Therapies

Genomic profiling technologies have allowed the stratification of human breast tumors into clinically useful groups and have further aided in the identification of targets for the treatment of breast cancer. The genomic era has produced an exponential increase in our understanding of cancer biology and has greatly accelerated

cancer drug development (**Figure 1.2**). With the advent and implementation of microarray expression profiling, it is now possible to evaluate gene expression in tumors on a genome-wide basis. These advances have led to the utilization of gene expression profiling to not only subtype cancers, but to predict prognosis and disease free survival, and determine optimal treatment [57-62].

Breast cancers that express the estrogen receptor-alpha account for 60-70% of breast cancers. In these cancers, estrogen plays a critical role in the etiology and progression of the disease. It is clear that the estrogen receptor and its ligand, estrogen, are critical for mitogenesis in this type of cancer. There remain, however, 30-40% of breast cancers that lack appreciable expression of estrogen receptor alpha and these tumors are poorly responsive to traditional therapies [63]. Regardless of the estrogen receptor status of the tumors, the growth signal sensed by the cell is conveyed to the nucleus through interactions of proteins in series, each one activating another, through signal transduction pathways. Once the signal is received in the nucleus, transcription factors activate genes important for cell growth and survival. Many of these pathways are understood in ER-positive cancers and have been the targets of small molecule inhibitors that can interrupt this mitogenic signaling, preventing and treating these cancers. Currently the mechanisms governing ER-negative breast cancer cell growth are largely unknown. It is clear that estrogen signaling is not the pathway that governs the mitogenic pathway in these ER-negative cells, however, for these ER-negative breast cancer cells, pathways critical for ER-negative growth have not been fully elucidated.

Figure 1.2



1940s

Mechlorethamine
Ethinly Estradiol

1950s

TEM
Mercaptopurine
Methotrexate
Chlorambucil
Cyclophosphamide
Thiotepa

1960s

Vinblastine
Uracil Mustard
Flurouracil
Vincristine
Melphalan
Actinomycin D
Pipobroman
Thioguanine
Hydroxyurea
Ara-C
Procarbazine

1970s

FUDR
Mithramycin
O-p'-DDD
Bleomycin
Adriamycin
Mitomycin C
Dacarbazine
CCNU
BCNU
Cis-Platinum
Daunomycin
Tamoxifen

1980s

Streptozotocin
Etoposide
Mitoxantrone
Ifosfamide
Carboplatin

1990s

Hexamethylmelamine
Idarubicin
Levamisole
Iudarabine Phophate
Pentostatin
Chorodeoxyadenosine
Taxol
Teniposide
Navelbine
All t-retinoic acid
Porfimer Na
Gemcitabine
Gliadel
Irinotecan
Taxotere
Topotecan
Herceptin
Ontak

Figure 1.2 Cancer Drug Development in the past half century: Graph shows cancer drugs receiving FDA approval in the past 50 years.

A major goal of current breast cancer research has been to identify targets that are unique to cancer cells and to identify drugs that kill only cancerous cells without affecting normal tissue. While achieving this goal has been difficult, emerging technologies have begun to shed light on pathways and molecules that may be involved in this malignant transformation. The use of gene expression profiling and other high-throughput technologies has allowed researchers to begin to untangle the complex cellular signaling networks and identify molecular “signatures” common to particular types of breast cancers.

1.10 Genomic, Transcriptional, and Proteomic Profiling of Breast Cancers:

Results from recent research have shown that breast cancer is a clinically heterogeneous disease. This clinical heterogeneity is driven to a large extent by abnormal gene expression within tumors. Investigators now have the ability to identify the gene-expression fingerprint of an individual's tumor. This information may be used to rationally design therapeutic targets in the future, and also to predict the clinical course of an individual's disease, including response to a specific treatment. Genetic profiles of tumors are now being correlated with clinical outcome, and several prognostic and predictive indicators have emerged based on this research. Additionally, transcriptional and proteomic profiling is advancing our understanding of the RNA and protein alterations in human cancers. Despite these early insights, it is clear that there is much still unknown, and the furthered utilization of these techniques will drive future

therapeutic development, lead to better risk stratification of patients, and guide rational therapy decisions by clinicians.

1.11 DNA-based genomic profiling:

DNA copy number determination using comparative genomic hybridization (CGH) or comparative single nucleotide polymorphism analysis (comparative SNP analysis) has shown that breast cancers harbor many gene deletions or gene amplification and that these regions of DNA copy number alteration identify genes or groups of genes that are involved in the oncogenic process [57-59, 61, 64]. Indeed, the well known breast cancer oncogenes HER2/*neu* and c-MYC, as well as more recently defined oncogenes *Rab25* [65], *NRG1* [62], and *LSM1* [63], have been identified using these DNA-based techniques. CGH and comparative SNP analysis also can identify regions of DNA loss, typically occurring at the site of important tumor suppressor genes. These techniques have identified the p53 and PTEN tumor suppressor genes specifically in breast cancer (both previously known tumor suppressor genes) [66, 67], as well as novel breast cancer tumor suppressor genes such as *PTK2b* [68] and *BRIT1* [69], and several other DNA regions in which tumor suppressor genes are thought to be located [70]. Indeed, with the increased effectiveness of DNA sequencing, groups have now begun to sequence large sets of human breast tumors and these efforts are leading to a furthered appreciation of the genomic aberrations harbored in human tumors. In a landmark study by Sjoblum *et al*., 13,023 genes in 11 breast and 11 colorectal cancers

were sequenced and subsequent analysis revealed that individual tumors accumulate an average of 90 mutant genes but that only a subset of these contribute to the neoplastic process [71]. They identified 189 genes (average of 11 per tumor) that were mutated at significant frequency, and that vast majority of these genes were not known to be genetically altered in tumors. These genes are predicted to affect a wide range of cellular functions, including transcription, adhesion, and invasion. These data define the genetic landscape of human breast tumors and provide new targets for diagnostic and therapeutic intervention. These studies have provided important insight into the chromosomal aberrations in human breast tumors, and may yet lead to the development of rational therapies targeting these aberrations.

1.12 RNA-based gene expression profiling to classify breast tumors and identify targets for the treatment of breast cancer.

It is now acknowledged that a continuum of abnormal gene expression predicts that tumorigenic phenotype and the sensitivity of tumors to treatment. Clinical investigators now have the capability to obtain a genetic blueprint of individual tumors; the genetic abnormalities identified within these tumors offer an opportunity to rationally select therapeutic targets for the treatment of patients with cancer [72]. Initial pioneering work in breast tumor classification came from Sorlie *et al*. who identified a gene set that was able to subdivide human breast tumors [73, 74]. They showed that ER-negative breast cancers could be classified into a basal epithelial-like

group, an *ERBB2*-overexpressing group and a normal breast-like group based on variations in gene expression. Additionally, they showed that the luminal epithelial/estrogen receptor-positive group could be divided into at least two subgroups, each with a distinctive expression profile (termed Luminal A and Luminal B). These subtypes proved to be reasonably robust by clustering using two different gene sets: first, a set of 456 cDNA clones chosen to reflect intrinsic properties of the tumors and, second, a gene set that highly correlated with patient outcome. Survival analyses on a subcohort of patients with locally advanced breast cancer uniformly treated in a prospective study showed significantly different outcomes for the patients belonging to the various groups, including a poor prognosis for the basal-like subtype and a significant difference in outcome for the two estrogen receptor-positive groups [73, 74]. This was the first of many studies to identify subtypes of human breast cancer based on gene expression profiling, and was the first to show that these subtypes had significantly different survival outcomes.

Ultimately, researchers aim to use the molecular data gathered from an individual tumor for prognostication and customization of therapy for each patient. Gene-expression profiling has shown promise to distinguish between patients at low and high risk for developing distant metastases and identify those who are likely to benefit from adjuvant therapy [75].

1.13 Gene-Expression Profiling and Prognostication.

The current standard for prognostic stratification includes Adjuvant! Online, the Nottingham Prognostic Index, and the American Joint Committee on Cancer staging system, which form the basis of treatment guidelines issued by the National Institutes of Health (NIH) Consensus Statement on Adjuvant Therapy in Breast Cancer and the St. Gallen Consensus Statement [76-79]. These tools integrate clinicopathologic factors into multivariate prediction models. Although these tools allow clinicians to estimate the relative risks for recurrence and mortality and estimate the potential benefits of chemotherapy for groups of patients with given disease characteristics [78], they do not address the fundamental question oncologists and patients struggle with: which individual (rather than which group) will benefit from adjuvant therapy? Up to 40%–50% of patients with a poor prognosis as defined by conventional clinicopathological parameters may remain disease free without adjuvant therapy [75]. Likewise, benefit from systemic adjuvant chemotherapy for patients with lymph node–negative (LNN) disease is not uniform; some patients relapse despite therapy and others may already be cured by local treatment. Therefore, a more individualized approach is needed. The goal of this thesis research is to define molecular abnormalities in individual tumors that can be exploited to treat ER-negative breast cancer more effectively.

1.14 Intrinsic Gene Set Identification

A novel molecular classification of breast cancer was proposed based on large-scale gene-expression analyses of breast cancer [71, 73, 74, 80-82]. From these studies, five major molecular classes of breast cancer emerged from several studies: luminal-A, luminal-B, basal-like, normal-like, and human epidermal growth factor receptor (HER)-positive cancers [73, 74, 80] (**Figure 1.3**). The Intrinsic subtype predictor was developed to assign molecular classes to newly diagnosed breast cancers [83]. These early profiling studies suggested that ER-negative tumors encompass three subgroups, one overexpressing ERBB2, one with tumors expressing genes characteristic of basal epithelial cells, and one with a gene expression profile similar to normal breast tissue. Clinically relevant novel subgroups within the ER-positive and ER-negative breast cancers have also been identified, reflecting the vastly different biology inherent in these tumor subtypes [73, 80]. ER-positive tumors were subdivided into a Luminal A and Luminal B subtype based on the gene expression differences inherently uniquely in these groups. This molecular taxonomy was based upon an "intrinsic" gene set, which was identified using a supervised analysis to select genes that showed little variance within repeated samplings of the same tumor, but which showed high variance across tumors [80]. These studies showed that an intrinsic gene set reflects the stable biological properties of tumors and typically identifies distinct tumor subtypes that have prognostic significance, even though no knowledge of outcome was used to derive this gene set. Despite the identification of multiple subtypes of breast cancer based on the

Figure 1.3

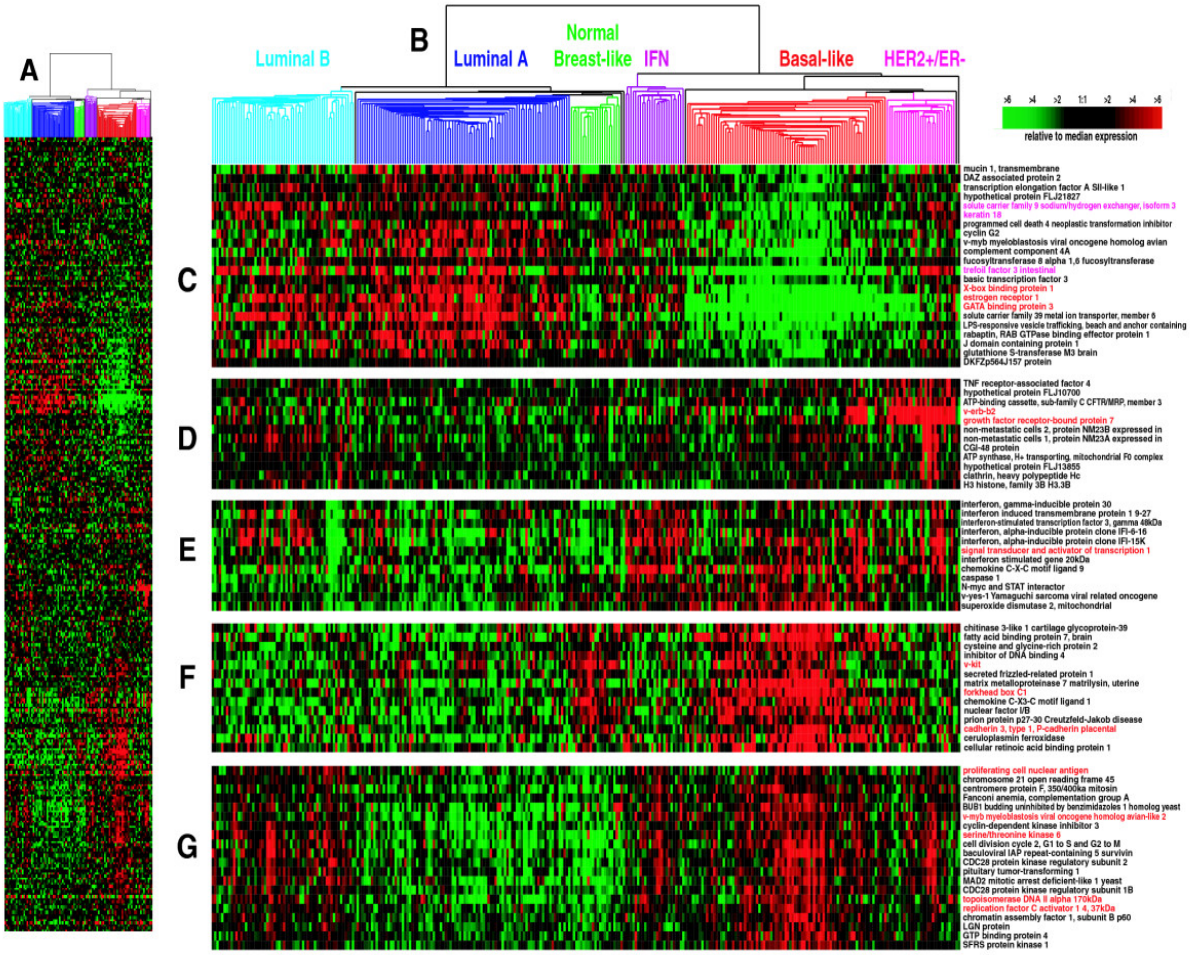


Figure 1.3. Hierarchical cluster analysis of the 315-sample combined test set using the Intrinsic/UNC gene set reduced to 306 genes. **(A)** Overview of complete cluster diagram. **(B)** Experimental sample-associated dendrogram. (**C**) Luminal/ER+ gene cluster with GATA3-regulated genes highlighted in pink. **(D)** HER2 and GRB7-containing expression cluster. **(E)** Interferon-regulated cluster containing STAT1. **(F)** Basal epithelial cluster. **(G)** Proliferation cluster. Adapted with permission from Hu Z, Chan F, Perou C *et al*. *BMC Genomics* 2006;7:96.

generation of an intrinsic stable gene list, a major challenge for microarray studies, especially those with clinical implications, has been validation of these gene lists [84, 85]. This limitation was addressed in a study by Hu *et al*. in 2006 [86], as well as others [73], that refined the intrinsic gene list that distinguishes these subtypes of human breast cancer. These analyses also showed that the breast tumor intrinsic subtypes are significant predictors of outcome when correcting for standard clinical parameters, and that common patterns of expression can be identified in data sets generated by independent labs.

The overall survival and chemotherapy sensitivity of the different molecular subgroups vary. Luminal-type cancers are mostly ER-positive, and patients with luminal-A cancers have the most favorable long-term survival (with endocrine therapy) compared with the other types, whereas basal-like and HER2–positive tumors, which are almost exclusively ER-negative, are more sensitive to chemotherapy [74, 80, 87]. These early studies not only identified an intrinsic gene list that has the ability to subtype breast tumors based on their gene expression profiles, but show that these differences in gene expression pattern may influence other characteristics including overall and metastasis free survival.

1.15 The Rotterdam Prognostic Gene Expression Signature

Multiple gene sets have been developed in an attempt to stratify patients based on the gene expression signature of their tumors. One of the first of these was the

Rotterdam gene set. It was developed to predict the prognosis of patients with lymph node negative (LNN) breast cancer [83]. Two hundred eighty-six patients who had locoregional therapy only were included in the initial development and validation study. Markers were selected separately from ER-negative and ER-positive tumors and were combined into a single 76-gene prognostic signature (VDX2; Veridex, LLC, Warren, NJ) that was able to predict distant metastatic recurrence with a sensitivity of 93% and a specificity of 48% [83]. This prognostic indicator performed better than standard, clinical variables in a multivariate analysis (hazard ratio [HR], 5.55; 95% confidence interval [CI], 2.46-12.5) [83]. Subsequently, this test was also validated using two other sets of patients with early stage breast cancer that were not included in the original study. The first set included 180 patients with stage I-II breast cancer and showed 5- and 10-year distant metastasis-free survival rates of 96% (95% CI, 89-99%) and 94% (95% CI, 83-98%), respectively, for the good prognosis group; the corresponding rates were 74% (95% CI, 64-81%) and 65% (95% CI, 53-74%) for the poor prognosis group [88]. The sensitivity for 5-year metastasis-free survival was 90%, and the specificity was 50%, with positive and negative predictive values of 38% and 94% respectively. The second validation cohort included 198 LNN cases and demonstrated similarly good 5- and 10-year distant metastasis-free survival rates: 98% (95% CI, 88-100%) and 94% (83-89%), respectively, for the genomic low-risk group [32]. The recurrence rates were significantly worse for the poor prognostic group: 76% (95% CI, 68-82%) and 73% (95 % CI, 65-79%) at 5 and 10 years, respectively [32]. Importantly, the 76-gene signature

could restratify patients within the clinical risk categories defined by the Adjuvant! Online program and the recurrence hazard ratios remained similar after adjustments for tumor grade, size, and ER status. This test in now FDA-approved and is clinically used to identify patients who should receive chemotherapy. The success of this gene expression profiling approach to address clinically relevant uncertainties underscores the utility of such profiling in the management of breast cancer

1.16 Other Gene Signatures

Other gene expression profiles have been used to characterize the different biological properties of breast cancers. One such signature, called the "Invasive gene signature" was identified based on the observation that low expression of CD24 and high expression of CD44 is highly tumorigenic in experimental models [89, 90]. Gene expression comparisons of this population of cells to normal breast epithelial cells identified 186 genes associated with the "tumorogenic breast stem cell" [90]. Similarly, tumors have been compared to nonhealing wounds [91]. Thus, a wound response indicator (WRI) was developed from genes whose expression changed following the activation of cultured fibroblasts with serum [91, 92]. This signature was applied to a cohort of tumors derived from patients with early-stage breast cancer that had gene expression profiling done (n = 295) [91]. Patients whose tumors expressed the WRI had significantly shorter overall survival and distant metastasis-free survival times relative to patients whose tumors did not express this gene signature. Moreover, the WRI

signature was an independent predictor of death in a multivariate analysis of metastasis and death. Again, both of these studies show that gene expression profiling is a powerful method for detecting potentially clinically relevant differences between types of breast cancer.

1.17 The onco*type* DX™ Recurrence Score™

RNA profiling is now used clinically to help identify individual tumors that will respond to chemotherapy. One such example is *oncotype* DX™ (Genomic Health, Inc; Redwood City, CA). The onco*type* DX™ is a 21-gene indicator. Two hundred fifty candidate genes were chosen from gene-expression profiling experiments, published literature, and genomic databases; these genes were correlated with breast cancer recurrence in 447 patients [93]. Sixteen cancer-related genes and five reference genes were selected from the candidate genes. The 16 cancer-related genes were then used to develop an algorithm based on the expression levels of these genes, thus allowing a Recurrence Score™ (RS) to be computed for each specimen. This RS correlated with the rate of distant recurrence at 10 years (**Figure 1.4**). This assay uses fixed tumor specimens, rather than frozen tissue. The onco*type* DX™ assay was externally validated in the National Surgical Adjuvant Breast and Bowel Project (NSABP) clinical trial B-14, which examined the effect of adjuvant tamoxifen in patients with hormone receptor–positive LNN breast cancer [93]. The results of this analysis showed that 7% of low-risk patients (RS <18) relapsed, whereas 31% of high-risk patients (RS >31) relapsed.

Figure 1.4

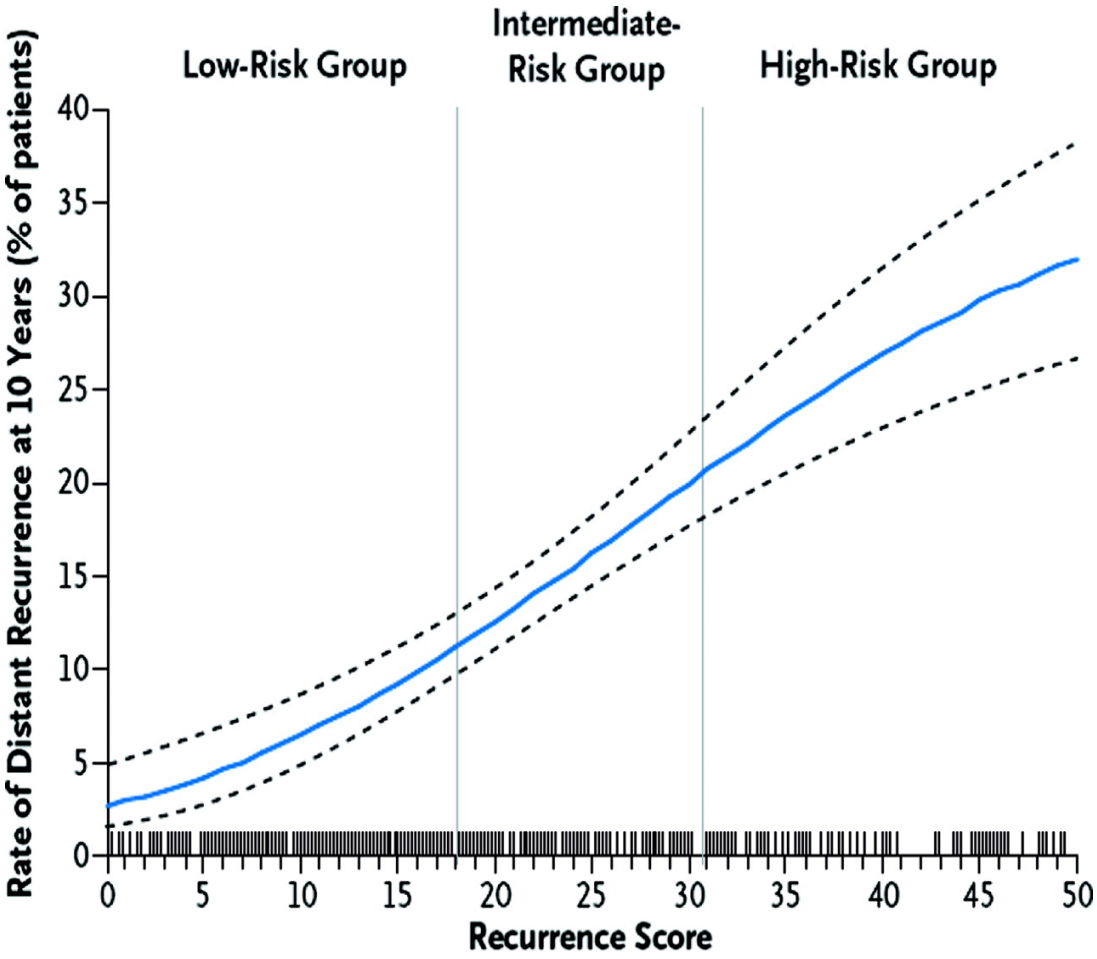


Figure 1.4. Rate of distant recurrence as a continuous function of the recurrence score. Reprinted with permission from Paik S, Shak S, Tang G et al, N Engl J Med 2004;351:2817–2826

Subsequent studies have shown that the RS is independently associated with sensitivity to chemotherapy and mortality [94, 95]. The onco*type* DX™ assay is now FDA-approved for use in profiling the risk and need for chemotherapy responsiveness in breast cancers. It is now in wide clinical use by oncologists to determine which patients would benefit from chemotherapy.

Onco*type* DX™ is not the only prognostic gene profiling test being used clinically. Other tests currently approved include Mammaprint [96], Mammostrat [97], and CellSearch [98]. These tests, which also mainly rely on the use of gene expression technologies and molecular signatures, underscore the power and utility of such approaches at identifying expression derangements and potential targets in breast cancer.

1.18 Utility of Genomic Profiling to Predict Response to Therapy:

In addition to defining molecular subtypes of cancers and predicting prognosis and disease free survival, gene expression microarray analysis has been used to determine optimal treatment [99-103]. Several groups have already used expression profiling to identify gene signatures of chemotherapeutic resistance [99, 100, 104]. These studies have identified tumor gene expression profiles associated with response to chemotherapy including docetaxel [99, 100] adriamycin/cyclophosphamide [82], paclitaxel, fluorouracil, doxorubicin, cyclophosphamide [105], and epirubicin, cyclophosphamide, paclitaxel [106] in the neoadjuvant setting. It has also been

demonstrated that PI3K pathway aberrations predict the likelihood of response to trastuzumab. In a complementary approach, identification of transcriptomes reflecting particular functional processes including expression profiles associated with ER-protein expression, histological grade, lymph node status, HER2/*neu* amplification, p53 mutation status, inflammatory breast cancer, and stromal signature patterns demonstrate the utility of the approach [73, 107-114]. Together, these studies indicate the potential to not only determine the likelihood of response to a particular therapy, but may be a means of identifying novel targets for therapy in ER-negative breast tumors.

1.19 Proteomic Profiling to Identify Tractable Targets in ER-negative Breast Cancer

Although gene expression technologies have proven very effective at subtyping breast tumors, they rely on the measurement of RNA molecules. These measurements, though useful, do not capture the expression and activation status of proteins, which are often the effector molecules in cells. Indeed, genomic and transcriptional information is ultimately translated into proteins that function to mediate cellular processes. Until recently, however, functional proteomics approaches have not been utilized to characterize the proteome of breast cancers.

Major efforts to discover new therapies for cancer have been brought about by advances in genomic and proteomic technologies and have resulted in many new potential drug targets. Most of these new targets are proteins involved in cellular

signaling, and the use of proteomic technologies provides new opportunities to measure the status of molecular networks as they exist within the context of the cellular milieu in both normal and diseased tissues. Protein microarrays can be used to profile the working state of cellular signal pathways in a manner not possible with gene microarrays since post-translational modifications cannot be accurately portrayed by global gene expression patterns alone [115-117]. Thus, unique opportunities exist for protein microarray technology, especially as it relates to therapeutic target discovery and validation. Using proteomic approaches it is possible to elucidate ongoing post-translational phosphorylation events and may even allow for the prediction of response in patients based upon the activity of the drug targets themselves. The technology may also be used to monitor total and phosphorylated proteins over time, before and after treatment, or between disease and non-disease states, allowing us to infer the activity levels of the proteins in a particular pathway in real time [118].

Recent advances now allow for proteomic arrays to subgroup breast cancers [119-125]. Using a novel quantitative protein detection system termed “reverse phase protein arrays” (RPPAs) which relies on validated high quality antibodies, expression levels and functional activation states of many signaling pathways can now be defined. This technique has provided a novel way to subclassify leukemias [123] and ovarian cancers [122]. RPPA is also able to quantitate very small amounts of protein expression (femtograms of target in nanograms of starting material), and in particular the activation state of cellular signaling pathways and networks using phospho-specific

antibodies. Thus, reverse phase protein arrays may be useful not only for target discovery, but also for validation, prediction of response, and identification of novel targets in ER-negative breast cancer that will usher in the age of truly personalized medicine.

1.20 Conclusions and Statement of Hypothesis.

Despite major advances in the treatment and prevention of breast cancer, this disease remains the second most common cause of cancer related death in women in the United States. Targeted treatment of estrogen receptor alpha-positive breast cancer using anti-estrogen drugs and aromatase inhibitors has resulted in a significant reduction in recurrence and mortality from this disease. Additionally, targeted treatment of women with HER2/*neu* receptor-positive breast cancer using treatments directed against the HER2/*neu* protein has reduced the recurrence of HER-positive breast cancers by 50%. Finally, other targeted therapies including growth factor receptor inhibitors, angiogenesis inhibitors, and signaling pathway inhibitors are currently being developed for the treatment of breast cancer. Unfortunately, many breast cancers do not express ER, HER2/*neu*, or these other growth factor receptors and no targeted therapy is available for these particularly aggressive cancers. These ER-negative tumors often arise in young women and are more aggressive and less responsive to therapy than ER-positive tumors. Thus, there is an urgent need for improved treatments for ER-negative breast cancer, particularly for tumors that lack the

estrogen receptor, the progesterone receptor, and the HER2/*neu* receptor ("triple-negative" breast cancer).

The overall goal of the research described herein was to use emerging genomic and proteomic technologies to identify novel drug targets for the treatment of ER-negative breast. We hypothesized that: **(1)** transcriptional profiling of human ER-negative breast cancers focused specifically on the kinome at the RNA level will identify kinases aberrantly expressed in ER-negative breast cancers, **(2)** that proteomic analysis will identify additional signaling molecules that are more highly expressed or activated in ER-negative breast cancers, and **(3)** that specific targeting of these aberrantly expressed molecules will inhibit breast cancer cell growth and will lead to the identification of novel targets for the effective treatment of human ER-negative breast cancer.

To address this hypothesis, three aims were developed:

Aim 1. To identify novel targets for the treatment of ER-negative breast cancer using gene expression profiling of kinases.

We first identified the kinases that were differentially expressed between ER-positive and ER-negative tumors. These results, described in **Chapter 3**, identify a distinct kinase gene expression profile that identifies ER-negative breast tumors and subsets ER-negative breast tumors into 4 distinct subgroups (cell cycle checkpoint, S6 kinase signaling, immunomodulatory, and MAPK signaling subgroups). Furthermore, we

have shown that the specific kinases overexpressed in ER-negative breast cancers are also overexpressed in ER-negative breast cancer cell lines and in an independent set of ER-negative human breast tumors. siRNA knockdown studies showed that several of these kinases are essential for the growth of ER-negative, but not ER-positive, breast cancer cell lines. Finally, survival analysis of patients with breast cancer demonstrated that the patients whose tumors had high expression of those kinases included in the S6 kinase signaling group of ER-negative cancer have an extremely poor prognosis, while patients whose tumors express high levels of immunomodulatory kinases have significantly better prognosis. These studies identify a list of kinases that are prognostic and may serve as druggable targets for the treatment of ER-negative breast cancer.

Aim 2. To identify novel targets for the treatment of ER-negative breast cancer using proteomic analyses.

In this aim, we identified differentially expressed proteins and phospho-proteins in ER-positive and ER-negative tumors to identify critical molecules and pathways that may be targeted for the effective treatment of ER-negative breast cancer. These results, described in **Chapter 4**, demonstrate that many proteins and pathways are elevated or activated in ER-negative breast cancer. Furthermore, it shows that ER-negative tumors can be subdivided into four distinct subgroups based on their expression of these proteins, and that these different subgroups have distinct prognostic profiles. It also identified protein signatures that are associated with particularly poor prognosis.

Finally, this chapter correlates specific proteomic signatures with previously described breast cancer subtypes identified by transcriptional profiling in human breast cancers. These studies identify proteins and pathways that are activated in specific subsets of ER-negative breast cancers that can now serve as targets of future drug development for effective treatment of ER-negative breast cancer.

Aim 3. Determine whether inhibition of one of the identified targets, maternal embryonic leucine zipper kinases (MELK), suppresses ER-negative breast cancer growth *in vitro.*

In this aim, we investigated whether specific inhibition of one of the targets indentified in the first aim was sufficient to inhibit ER-negative breast cancer growth and invasion. The results described in **Chapter 5** demonstrate that MELK is more highly expressed in ER-negative breast cancers. We confirm that MELK RNA and protein are more highly expressed in high grade and undifferentiated tumors and breast cancer cell lines. Further we show that specific knockdown of MELK using siRNA technologies inhibits the growth of ER-negative breast cancer cells, but not ER-positive breast cancer cells. Finally, we show that MELK expression is correlated with poor metastasis-free and overall survival in multiple datasets, and is independently prognostic (independent of grade, tumor size, age, ER-status, and lymph node status) in breast cancer; tumors with high MELK expression have a very poor outcome. These results provide the pre-clinical

rationale for the development of MELK inhibitors for the treatment of ER-negative breast cancer.

References:

1. *American Cancer Society, Cancer Facts and Figures 2007 (American Cancer Society, Atlanta, Georgia), 2007*.
2. Carey, L.A., et al., *Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.* Jama, 2006. **295**(21): p. 2492-502.
3. Couse, J.F., et al., *Tissue distribution and quantitative analysis of estrogen receptor-alpha (ERalpha) and estrogen receptor-beta (ERbeta) messenger ribonucleic acid in the wild-type and ERalpha-knockout mouse.* Endocrinology, 1997. **138**(11): p. 4613-21.
4. Patel, R.R., C.G. Sharma, and V.C. Jordan, *Optimizing the antihormonal treatment and prevention of breast cancer.* Breast Cancer, 2007. **14**(2): p. 113-22.
5. Howell, A., et al., *Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer.* Lancet, 2005. **365**(9453): p. 60-2.
6. Thurlimann, B., et al., *A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer.* N Engl J Med, 2005. **353**(26): p. 2747-57.
7. Fisher, B., et al., *Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study.* J Natl Cancer Inst, 2005. **97**(22): p. 1652-62.
8. Cummings, S.R., et al., *The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation.* Jama, 1999. **281**(23): p. 2189-97.
9. Geyer, C.E., et al., *Lapatinib plus capecitabine for HER2-positive advanced breast cancer.* N Engl J Med, 2006. **355**(26): p. 2733-43.
10. Moy, B., et al., *Lapatinib.* Nat Rev Drug Discov, 2007. **6**(6): p. 431-2.
11. Mayer, E.L., N.U. Lin, and H.J. Burstein, *Novel approaches to advanced breast cancer: bevacizumab and lapatinib.* J Natl Compr Canc Netw, 2007. **5**(3): p. 314-23.
12. Sonpavde, G., *Lapatinib plus capecitabine in breast cancer.* N Engl J Med, 2007. **356**(14): p. 1471; author reply 1471-2.
13. Dawood, S., et al., *Efficacy and safety of neoadjuvant trastuzumab combined with paclitaxel and epirubicin: a retrospective review of the M. D. Anderson experience.* Cancer, 2007. **110**(6): p. 1195-200.
14. Burstein, H.J., et al., *Trastuzumab plus vinorelbine or taxane chemotherapy for HER2-overexpressing metastatic breast cancer: the trastuzumab and vinorelbine or taxane study.* Cancer, 2007. **110**(5): p. 965-72.
15. Garrison, L.P., Jr., et al., *Cost-effectiveness analysis of trastuzumab in the adjuvant setting for treatment of HER2-positive breast cancer.* Cancer, 2007. **110**(3): p. 489-98.

16. Berns, K., et al., *A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer.* Cancer Cell, 2007. **12**(4): p. 395-402.
17. Linggi, B. and G. Carpenter, *ErbB receptors: new insights on mechanisms and biology.* Trends Cell Biol, 2006. **16**(12): p. 649-56.
18. Slamon, D.J., et al., *Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2.* N Engl J Med, 2001. **344**(11): p. 783-92.
19. Cobleigh, M.A., et al., *Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease.* J Clin Oncol, 1999. **17**(9): p. 2639-48.
20. Vogel, C.L., et al., *Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer.* J Clin Oncol, 2002. **20**(3): p. 719-26.
21. Hussain, M.H., et al., *Trastuzumab, paclitaxel, carboplatin, and gemcitabine in advanced human epidermal growth factor receptor-2/neu-positive urothelial carcinoma: results of a multicenter phase II National Cancer Institute trial.* J Clin Oncol, 2007. **25**(16): p. 2218-24.
22. Smith, I., et al., *2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial.* Lancet, 2007. **369**(9555): p. 29-36.
23. Nahta, R. and F.J. Esteva, *HER2 therapy: molecular mechanisms of trastuzumab resistance.* Breast Cancer Res, 2006. **8**(6): p. 215.
24. Nahta, R., et al., *Mechanisms of disease: understanding resistance to HER2-targeted therapy in human breast cancer.* Nat Clin Pract Oncol, 2006. **3**(5): p. 269-80.
25. Konecny, G.E., et al., *Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells.* Cancer Res, 2006. **66**(3): p. 1630-9.
26. Tuma, R.S., *Lapatinib moves forward in inflammatory and early HER2-positive breast cancer trials.* J Natl Cancer Inst, 2007. **99**(5): p. 348-9.
27. Toi, M., et al., *Association of vascular endothelial growth factor expression with tumor angiogenesis and with early relapse in primary breast cancer.* Jpn J Cancer Res, 1994. **85**(10): p. 1045-9.
28. Yamaguchi, T., et al., *Overexpression of soluble vascular endothelial growth factor receptor 1 in colorectal cancer: Association with progression and prognosis.* Cancer Sci, 2007. **98**(3): p. 405-10.
29. Miller, K.D., et al., *Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer.* J Clin Oncol, 2005. **23**(4): p. 792-9.

30. Cobleigh, M.A., et al., *A phase I/II dose-escalation trial of bevacizumab in previously treated metastatic breast cancer.* Semin Oncol, 2003. **30**(5 Suppl 16): p. 117-24.
31. Schneider, B.P., et al., *Association of vascular endothelial growth factor and vascular endothelial growth factor receptor-2 genetic polymorphisms with outcome in a trial of paclitaxel compared with paclitaxel plus bevacizumab in advanced breast cancer: ECOG 2100.* J Clin Oncol, 2008. **26**(28): p. 4672-8.
32. Miller, K., et al., *Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer.* N Engl J Med, 2007. **357**(26): p. 2666-76.
33. Miller, K.D., *E2100: a phase III trial of paclitaxel versus paclitaxel/bevacizumab for metastatic breast cancer.* Clin Breast Cancer, 2003. **3**(6): p. 421-2.
34. Dickler, M.N., et al., *A phase II trial of erlotinib in combination with bevacizumab in patients with metastatic breast cancer.* Clin Cancer Res, 2008. **14**(23): p. 7878-83.
35. Sell, C., et al., *Simian virus 40 large tumor antigen is unable to transform mouse embryonic fibroblasts lacking type 1 insulin-like growth factor receptor.* Proc Natl Acad Sci U S A, 1993. **90**(23): p. 11217-21.
36. Carboni, J.M., et al., *Tumor development by transgenic expression of a constitutively active insulin-like growth factor I receptor.* Cancer Res, 2005. **65**(9): p. 3781-7.
37. Yakar, S., D. Leroith, and P. Brodt, *The role of the growth hormone/insulin-like growth factor axis in tumor growth and progression: Lessons from animal models.* Cytokine Growth Factor Rev, 2005. **16**(4-5): p. 407-20.
38. Nielsen, T.O., et al., *Expression of the insulin-like growth factor I receptor and urokinase plasminogen activator in breast cancer is associated with poor survival: potential for intervention with 17-allylamino geldanamycin.* Cancer Res, 2004. **64**(1): p. 286-91.
39. Bonneterre, J., et al., *Prognostic significance of insulin-like growth factor 1 receptors in human breast cancer.* Cancer Res, 1990. **50**(21): p. 6931-5.
40. Mitsiades, C.S., et al., *Inhibition of the insulin-like growth factor receptor-1 tyrosine kinase activity as a therapeutic strategy for multiple myeloma, other hematologic malignancies, and solid tumors.* Cancer Cell, 2004. **5**(3): p. 221-30.
41. Haluska, P., et al., *In vitro and in vivo antitumor effects of the dual insulin-like growth factor-I/insulin receptor inhibitor, BMS-554417.* Cancer Res, 2006. **66**(1): p. 362-71.
42. Wittman, M., et al., *Discovery of a (1H-benzoimidazol-2-yl)-1H-pyridin-2-one (BMS-536924) inhibitor of insulin-like growth factor I receptor kinase with in vivo antitumor activity.* J Med Chem, 2005. **48**(18): p. 5639-43.
43. Lee, A.V., X. Cui, and S. Oesterreich, *Cross-talk among estrogen receptor, epidermal growth factor, and insulin-like growth factor signaling in breast*

cancer. Clin Cancer Res, 2001. **7**(12 Suppl): p. 4429s-4435s; discussion 4411s-4412s.
44. Liang, K., et al., *Targeting the phosphatidylinositol 3-kinase/Akt pathway for enhancing breast cancer cells to radiotherapy.* Mol Cancer Ther, 2003. **2**(4): p. 353-60.
45. Campbell, R.A., et al., *Phosphatidylinositol 3-kinase/AKT-mediated activation of estrogen receptor alpha: a new model for anti-estrogen resistance.* J Biol Chem, 2001. **276**(13): p. 9817-24.
46. Mondesire, W.H., et al., *Targeting mammalian target of rapamycin synergistically enhances chemotherapy-induced cytotoxicity in breast cancer cells.* Clin Cancer Res, 2004. **10**(20): p. 7031-42.
47. Pandolfi, P.P., *Breast cancer--loss of PTEN predicts resistance to treatment.* N Engl J Med, 2004. **351**(22): p. 2337-8.
48. Bhandare, N., et al., *Ototoxicity after radiotherapy for head and neck tumors.* Int J Radiat Oncol Biol Phys, 2007. **67**(2): p. 469-79.
49. deGraffenried, L.A., et al., *Inhibition of mTOR activity restores tamoxifen response in breast cancer cells with aberrant Akt Activity.* Clin Cancer Res, 2004. **10**(23): p. 8059-67.
50. DeGraffenried, L.A., et al., *Reduced PTEN expression in breast cancer cells confers susceptibility to inhibitors of the PI3 kinase/Akt pathway.* Ann Oncol, 2004. **15**(10): p. 1510-6.
51. Punt, C.J., et al., *Phase I and pharmacokinetic study of CCI-779, a novel cytostatic cell-cycle inhibitor, in combination with 5-fluorouracil and leucovorin in patients with advanced solid tumors.* Ann Oncol, 2003. **14**(6): p. 931-7.
52. Baselga J, F.O., Gil M, Colomer R, Roche H, Cortes-Funes H, et al. , *A phase II 3-arm study of CCI-779 in combination with letrozole in postmenopausal women with locally advanced or metastatic breast cancer.* Proc Am Soc Clin Oncol, 2004. **23**(544).
53. Hennessy, B.T., et al., *Exploiting the PI3K/AKT pathway for cancer drug discovery.* Nat Rev Drug Discov, 2005. **4**(12): p. 988-1004.
54. Stemke-Hale, K., et al., *An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer.* Cancer Res, 2008. **68**(15): p. 6084-91.
55. Murph, M.M., et al., *Individualized molecular medicine: linking functional proteomics to select therapeutics targeting the PI3K pathway for specific patients.* Adv Exp Med Biol, 2008. **622**: p. 183-95.
56. Tan, A.R. and S.M. Swain, *Adjuvant chemotherapy for breast cancer: an update.* Semin Oncol, 2001. **28**(4): p. 359-76.
57. Pollack, J.R., et al., *Genome-wide analysis of DNA copy-number changes using cDNA microarrays.* Nat Genet, 1999. **23**(1): p. 41-6.
58. Rodriguez, C., et al., *At least five regions of imbalance on 6q in breast tumors, combining losses and gains.* Genes Chromosomes Cancer, 2000. **27**(1): p. 76-84.

59. Werner, M., et al., *20q13.2 amplification in intraductal hyperplasia adjacent to in situ and invasive ductal carcinoma of the breast.* Virchows Arch, 1999. **435**(5): p. 469-72.
60. Johnson, N., et al., *A comparative study of genome-wide SNP, CGH microarray and protein expression analysis to explore genotypic and phenotypic mechanisms of acquired antiestrogen resistance in breast cancer.* Breast Cancer Res Treat, 2008. **111**(1): p. 55-63.
61. Climent, J., et al., *Characterization of breast cancer by array comparative genomic hybridization.* Biochem Cell Biol, 2007. **85**(4): p. 497-508.
62. Prentice, L.M., et al., *NRG1 gene rearrangements in clinical breast cancer: identification of an adjacent novel amplicon associated with poor prognosis.* Oncogene, 2005. **24**(49): p. 7281-9.
63. Yang, Z.Q., et al., *Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer.* Cancer Res, 2006. **66**(24): p. 11632-43.
64. Johnson, N., et al., *A comparative study of genome-wide SNP, CGH microarray and protein expression analysis to explore genotypic and phenotypic mechanisms of acquired antiestrogen resistance in breast cancer.* Breast Cancer Res Treat, 2007.
65. Cheng, K.W., et al., *The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers.* Nat Med, 2004. **10**(11): p. 1251-6.
66. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.
67. Wong, K.K., et al., *Allelic imbalance analysis by high-density single-nucleotide polymorphic allele (SNP) array with whole genome amplified DNA.* Nucleic Acids Res, 2004. **32**(9): p. e69.
68. Naylor, T.L., et al., *High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization.* Breast Cancer Res, 2005. **7**(6): p. R1186-98.
69. Rai, R., et al., *BRIT1 regulates early DNA damage response, chromosomal integrity, and cancer.* Cancer Cell, 2006. **10**(2): p. 145-57.
70. Fridlyand, J., et al., *Breast tumor copy number aberration phenotypes and genomic instability.* BMC Cancer, 2006. **6**: p. 96.
71. Sjoblom, T., et al., *The consensus coding sequences of human breast and colorectal cancers.* Science, 2006. **314**(5797): p. 268-74.
72. Morris, S.R. and L.A. Carey, *Molecular profiling in breast cancer.* Rev Endocr Metab Disord, 2007. **8**(3): p. 185-98.
73. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
74. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.

75. van de Vijver, M., *Gene-expression profiling and the future of adjuvant therapy.* Oncologist, 2005. **10 Suppl 2**: p. 30-4.
76. Goldhirsch, A., et al., *Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005.* Ann Oncol, 2005. **16**(10): p. 1569-83.
77. D'Eredita, G., et al., *Prognostic factors in breast cancer: the predictive value of the Nottingham Prognostic Index in patients with a long-term follow-up that were treated in a single institution.* Eur J Cancer, 2001. **37**(5): p. 591-6.
78. Olivotto, I.A., et al., *Population-based validation of the prognostic model ADJUVANT! for early breast cancer.* J Clin Oncol, 2005. **23**(12): p. 2716-25.
79. Singletary, S.E., et al., *Staging system for breast cancer: revisions for the 6th edition of the AJCC Cancer Staging Manual.* Surg Clin North Am, 2003. **83**(4): p. 803-19.
80. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.
81. Sotiriou, C., et al., *Breast cancer classification and prognosis based on gene expression profiles from a population-based study.* Proc Natl Acad Sci U S A, 2003. **100**(18): p. 10393-8.
82. Sotiriou, C., et al., *Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer.* Breast Cancer Res, 2002. **4**(3): p. R3.
83. Wang, Y., et al., *Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.* Lancet, 2005. **365**(9460): p. 671-9.
84. Michiels, S., S. Koscielny, and C. Hill, *Prediction of cancer outcome with microarrays: a multiple random validation strategy.* Lancet, 2005. **365**(9458): p. 488-92.
85. Jenssen, T.K. and E. Hovig, *Gene-expression profiling in breast cancer.* Lancet, 2005. **365**(9460): p. 634-5.
86. Hu, Z., et al., *The molecular portraits of breast tumors are conserved across microarray platforms.* BMC Genomics, 2006. **7**: p. 96.
87. Rouzier, R., et al., *Breast cancer molecular subtypes respond differently to preoperative chemotherapy.* Clin Cancer Res, 2005. **11**(16): p. 5678-85.
88. Foekens, J.A., et al., *Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer.* J Clin Oncol, 2006. **24**(11): p. 1665-71.
89. Al-Hajj, M., et al., *Prospective identification of tumorigenic breast cancer cells.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3983-8.
90. Liu, R., et al., *The prognostic role of a gene signature from tumorigenic breast-cancer cells.* N Engl J Med, 2007. **356**(3): p. 217-26.

91. Chang, H.Y., et al., *Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival.* Proc Natl Acad Sci U S A, 2005. **102**(10): p. 3738-43.
92. Chang, H.Y., et al., *Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds.* PLoS Biol, 2004. **2**(2): p. E7.
93. Paik, S., et al., *A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.* N Engl J Med, 2004. **351**(27): p. 2817-26.
94. Gianni, L., et al., *Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer.* J Clin Oncol, 2005. **23**(29): p. 7265-77.
95. Paik, S., et al., *Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer.* J Clin Oncol, 2006. **24**(23): p. 3726-34.
96. Buyse, M., et al., *Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer.* J Natl Cancer Inst, 2006. **98**(17): p. 1183-92.
97. Ring, B.Z., et al., *Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer.* J Clin Oncol, 2006. **24**(19): p. 3039-47.
98. Cristofanilli, M., et al., *Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer.* J Clin Oncol, 2005. **23**(7): p. 1420-30.
99. Chang, J.C., et al., *Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer.* Lancet, 2003. **362**(9381): p. 362-9.
100. Chang, J.C., et al., *Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients.* J Clin Oncol, 2005. **23**(6): p. 1169-77.
101. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.
102. van 't Veer, L.J., et al., *Expression profiling predicts outcome in breast cancer.* Breast Cancer Res, 2003. **5**(1): p. 57-8.
103. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-6.
104. Chang, J., et al., *Survival of patients with metastatic breast carcinoma: importance of prognostic markers of the primary tumor.* Cancer, 2003. **97**(3): p. 545-53.
105. Ayers, M., et al., *Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer.* J Clin Oncol, 2004. **22**(12): p. 2284-93.

106. Modlich, O., et al., *Immediate gene expression changes after the first course of neoadjuvant chemotherapy in patients with primary breast cancer disease.* Clin Cancer Res, 2004. **10**(19): p. 6418-31.
107. Gruvberger, S., et al., *Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns.* Cancer Res, 2001. **61**(16): p. 5979-84.
108. West, M., et al., *Predicting the clinical status of human breast cancer by using gene expression profiles.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11462-7.
109. Hedenfalk, I., et al., *Molecular classification of familial non-BRCA1/BRCA2 breast cancer.* Proc Natl Acad Sci U S A, 2003. **100**(5): p. 2532-7.
110. Ahr, A., et al., *Molecular classification of breast cancer patients by gene expression profiling.* J Pathol, 2001. **195**(3): p. 312-20.
111. Iwao, K., et al., *Molecular classification of primary breast tumors possessing distinct prognostic properties.* Hum Mol Genet, 2002. **11**(2): p. 199-206.
112. Dressman, M.A., et al., *Gene expression profiling detects gene amplification and differentiates tumor types in breast cancer.* Cancer Res, 2003. **63**(9): p. 2194-9.
113. Bieche, I., et al., *Molecular profiling of inflammatory breast cancer: identification of a poor-prognosis gene expression signature.* Clin Cancer Res, 2004. **10**(20): p. 6789-95.
114. Miller, L.D., et al., *An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival.* Proc Natl Acad Sci U S A, 2005. **102**(38): p. 13550-5.
115. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-65.
116. MacBeath, G., *Protein microarrays and proteomics.* Nat Genet, 2002. **32 Suppl**: p. 526-32.
117. Wilson, D.S. and S. Nock, *Recent developments in protein microarray technology.* Angew Chem Int Ed Engl, 2003. **42**(5): p. 494-500.
118. Carr, K.M., et al., *Genomic and proteomic approaches for studying human cancer: prospects for true patient-tailored therapy.* Hum Genomics, 2004. **1**(2): p. 134-40.
119. Kattenhorn, L.M., et al., *Identification of proteins associated with murine cytomegalovirus virions.* J Virol, 2004. **78**(20): p. 11187-97.
120. Petricoin, E.F., et al., *Use of proteomic patterns in serum to identify ovarian cancer.* Lancet, 2002. **359**(9306): p. 572-7.
121. Polley, A.C., et al., *Proteomic analysis reveals field-wide changes in protein expression in the morphologically normal mucosa of patients with colorectal neoplasia.* Cancer Res, 2006. **66**(13): p. 6553-62.
122. Sheehan, K.M., et al., *Use of reverse phase protein microarrays and reference standard development for molecular network analysis of metastatic ovarian carcinoma.* Mol Cell Proteomics, 2005. **4**(4): p. 346-55.

123. Tibes, R., et al., *Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells.* Mol Cancer Ther, 2006. **5**(10): p. 2512-21.
124. Nimeus, E., et al., *Proteomic analysis identifies candidate proteins associated with distant recurrences in breast cancer after adjuvant chemotherapy.* J Pharm Biomed Anal, 2007. **43**(3): p. 1086-93.
125. Stemke-Hale, K., et al., *Molecular screening for breast cancer prevention, early detection, and treatment planning: combining biomarkers from DNA, RNA, and protein.* Curr Oncol Rep, 2006. **8**(6): p. 484-91.

Chapter 2

Material and Methods

2.1 Introduction

A number of materials and methods were used in the studies that comprise this thesis. To fully describe the details of these methods, this chapter is divided into two sections. The first section, "General Material and Methods" broadly describes materials and methods that were used in experiments in multiple chapters. The second section, "Specific Materials and Methods", details the methods used only in the experiments contained in that particular chapter. All methods, analysis, and experimentation was done by Corey Speers unless otherwise noted. Statistical supervision and guidance was provided by Dr. Susan Hilsenbeck at Baylor College of Medicine and Dr. John Quackenbush at the Dana Farber Cancer Institute.

2.2 General Materials and Methods

The techniques described herein are those techniques common to experimentation throughout this thesis. The technical details of these methods are listed, however the experiment and project-specific conditions are listed in the accompanying section, "Specific Materials and Methods".

2.2.1 Cell Lines and Cell culture:

MCF-7, T47D, BT474, MDA-MB-361, Hs578T, MDA-MB-231, BT549, MDA-MB-468, HCC1937, HCC1569, HCC1187, SKBr3, MDA-MB-453, BT20, and ZR75-1 cells were maintained as outlined in **Table 2.1**. Also listed in **Table 2.1** are the characteristics of

Table 2.1 - Cell lines used in experiments

Cell line	Gene cluster	ER	PR	HER2	TP53	Source	Tumor type	Age (years)	Ethnicity	Culture media	Culture conditions
AU565a	Lu	−	[−]	+	$+^{WT}$	PE	AC	43	W	RPMI, 10% FBS	37°C, 5% CO_2
BT20	BaA	−	[−]		$++^{WT}$	P.Br	IDC	74	W	DMEM, 10% FBS	37°C, 5% CO_2
BT474	Lu	+	[+]	+	+	P.Br	IDC	60	W	RPMI, 10% FBS	37°C, 5% CO_2
BT549	BaB	−	[−]		$++^{M}$	P.Br	IDC, pap	72	W	RPMI, 10% FBS	37°C, 5% CO_2
HCC1007	Lu	+	[−]		[+/−]	P.Br	Duc.Ca	67	B	RPMI, 10% FBS	37°C, 5% CO_2
HCC1187	BaA	−	[−]		$++^{M}$	P.Br	Duc.Ca	41	W	RPMI, 10% FBS	37°C, 5% CO_2
HCC1569	BaA	−	[−]	+	$-^{M}$	P.Br	MC	70	B	RPMI, 10% FBS	37°C, 5% CO_2
HCC1937	BaA	−	[−]		[−]	P.Br	Duc.Ca	24	W	RPMI, 10% FBS	37°C, 5% CO_2
HCC1954	BaA	−	[−]	+	[+/−]	P.Br	Duc.Ca	61	EI	RPMI, 10% FBS	37°C, 5% CO_2
HCC202	Lu	−	[−]	+	[−]	P.Br	Duc.Ca	82	W	RPMI, 10% FBS	37°C, 5% CO_2
HCC38	BaB	−	[−]		$++^{M}$	P.Br	Duc.Ca	50	W	RPMI, 10% FBS	37°C, 5% CO_2
HCC70	BaA	−	[−]		$++^{M}$	P.Br	Duc.Ca	49	B	RPMI, 10% FBS	37°C, 5% CO_2
HS578T	BaB	−	[−]		$+^{M}$	P.Br	IDC	74	W	DMEM, 10% FBS	37°C, 5% CO_2
LY2	Lu	+	[−]		+/−	PE	IDC	69	W	DMEM, 10% FBS	37°C, 5% CO_2
MCF10A	BaB	−	[−]		$+/-^{WT}$	P.Br	F	36	W	DMEM/F12*	37°C, 5% CO_2
MCF12A	BaB	−	[−]		+	P.Br	F	60	W	DMEM/F12*	37°C, 5% CO_2
MCF7	Lu	+	[+]		$+/-^{WT}$	PE	IDC	69	W	DMEM, 10% FBS	37°C, 5% CO_2
MDAMB231	BaB	−	[−]		$++^{M}$	PE	AC	51	W	DMEM, 10% FBS	37°C, 5% CO_2
MDAMB361	Lu	+	[−]	+	$-^{WT}$	P.Br	AC	40	W	DMEM, 10% FBS	37°C, 5% CO_2
MDAMB435	BaB	−	[−]		$+^{M}$	PE	IDC	31	W	DMEM, 10% FBS	37°C, 5% CO_2

MDAMB453	Lu	–	[–]		$-^{WT}$	PF	AC	48	W	DMEM, 10% FBS	37°C, 5% CO_2
MDAMB468	BaA	[–]	[–]		[+]	PE	AC	51	B	L15, 10% FBS	37°C, no CO_2
SKBR3	Lu	–	[–]	+	+	PE	AC	43	W	DMEM, 10% FBS	37°C, 5% CO_2
SUM149PT	BaB	[–]	[–]		[+]	P.Br	Inf Duc.Ca	UN	UN	Ham's F12, 5%-IH	37°C, 5% CO_2
SUM185PE	Lu	[–]	[–]		[–]	PE	Duc.Ca	UN	UN	Ham's F12, 5%-IH	37°C, 5% CO_2
SUM190PT	BaA	–	[–]	+	[+/–]	P.Br	Inf	UN	UN	Ham's F12, SF-IH	37°C, 5% CO_2
T47D	Lu	+	[+]		$++^{M}$	PE	IDC	54	UN	RPMI, 10% FBS	37°C, 5% CO_2
ZR75B	Lu	+	[–]		+/–	UN	UN	UN	UN	RPMI, 10% FBS	37°C, 5% CO_2

AC, adenocarcinoma; AF, ascites fluid; AnCa, anaplastic carcinoma; ASC, acantholytic squamous carcinoma;
BaA, Basal A; BaB, Basal B; Ca, carcinoma; CWN, chest wall nodule; Duc.Ca, ductal carcinoma; F, fibrocystic disease;
IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; Inf, inflammatory; LN, lymph node; Lu, luminal;
MC, metaplastic carcinoma; MLCa, metastatic lobular carcinoma; N, normal; Pap, papillary; ND, not done;
P.Br, primary breast; PE, pleural effusion; Sk, skin; W, White; B, Black; H, Hispanic; EI; East Indian.
UN, unknown

Table 2.1 - A list of cell lines and their characterstics for all cell lines used in these studies

the cell lines used in the experiments throughout this thesis. Generally, cells were maintained in Improved MEM (IMEM) Zn++ Option (Richter's Modification, Invitrogen, Carlsbad, CA) with either 5% or 10% heat inactivated fetal bovine serum (FBS, Hyclone, Logan, UT) and 1% Penicillin-Streptomycin-Glutamine (Invitrogen) supplemented with 5% CO_2.

2.2.2 Western blot analysis:

Cells were washed once with ice-cold phosphate buffered saline (PBS) and lysed in protein lysis buffer consisting of 50mM HEPES pH7.5, 150mM NaCl, 1mM EDTA, 1% Triton X-100, 10% glycerol, 100mM NaF, Complete Mini protease inhibitors cocktail tablet (Roche), and phosphatase inhibitor cocktail I and II (Sigma-Aldrich). Protein concentration was determined using BCA Protein Assay Reagents (Pierce Biotechnology, Rockford, IL). An aliquot of total protein (20μg) was resolved by electrophoresis in 10% SDS-PAGE gel and transferred to a PVDF membrane (Millipore, Billerica, MA). The membrane was blocked with 5% nonfat milk in 10mM Tris-HCl pH7.4, 150mM NaCl, and 0.1% Tween 20 (TBST) overnight at 4°C. Thereafter, the membrane was incubated with primary antibody diluted in 1% nonfat milk/TBST overnight at 4°C or 3 hours at room temperature, after which the membrane was washed in TBST 3 times for 10 min each. The membrane was then incubated with horseradish peroxidase-conjugated secondary antibody diluted in 1% milk/TBST at room temperature for 1 hour. The membrane was washed 3 times in TBST for 10 minutes each. Antigen-antibody complexes were

detected using the ECL or ECL Plus chemiluminescent system (Amersham Bioscience, Piscataway, NJ). Antibodies specific for beta-actin (diluted 1:10,000, #A-5441) were purchased from Sigma-Aldrich. Anti-mouse (diluted 1:5000, #NA931V) and anti-rabbit (diluted 1:5000, #NA934V) secondary antibodies were obtained from Amersham Bioscience (Piscataway, NJ).

2.2.3 RNA preparation and quantitative RT-PCR:

Total RNA was isolated using the RNeasy RNA isolation kit (QIAGEN, Valencia, CA). Briefly, cells were lysed in cell lysis buffer using a cell lifter or a rotor stator homogenizer (Pro Scientific, Oxford, CT). Ethanol was added to the lysate creating conditions that promote selective binding of the RNA to the silica-based RNA binding column. After adding lysate that includes RNA to the silica column, washing buffers were used to eliminate unwanted associated proteins and DNA while retaining the RNA on the column. Ethanol was again used to remove contaminants and the RNA was eluted off the column with elution buffer. RNA concentration and quality was assessed on a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE). Quantitative RT-PCR (Q-RT-PCR) assays of transcripts were carried out using gene-specific double fluorescence-labeled probes in an ABI PRISM 7700 Fast Sequence Detector (Applied Biosystem, Foster City, CA). All primers, probes, and standards used in these studies were designed using the universal mouse and human probe set library from Roche and cDNA was generated using oligo dT or gene specific primers. All assays were designed

using annotated mRNA sequences from the Entrez NCBI site. The PCR reaction mixture consisted of 300nM each of the forward and reverse primers, 100nM probe, 0.025 units/µl of Taq Polymerase (Invitrogen), 125µM each of dNTP, 5mM MgCl2, and 1X Taq Polymerase buffer. Cycling conditions were 94°C for 30 seconds, followed by 40 cycles at 94°C for 10 seconds and 60°C for 5 seconds. 6-Carboxy fluorescein (FAM) was used as the 5' fluorescent reporter and black hole quencher (BHQ1) was used at the 3' end quencher. All reactions were performed using at least triplicate RNA samples independently isolated. Q-RT-PCR assays were performed in triplicate for each sample. Standard curves for the quantification of each transcript were generated using the serially diluted solution of synthetic templates. The number of transcript molecules was calculated by extrapolation using this standard curve. Data were reported as average number of molecules ± standard error of the mean.

2.2.4 siRNA transfection:

siRNAs for all genes of interest were purchased either from Dharmacon Research (Lafayette, CO) or Sigma Aldrich (St. Louis, MO). siRNA transfection was performed using DharmaFECT^{TM} 1 (Dharmacon), according to the manufacture's instruction into breast cancer cell lines. Briefly, cells were grown to 60-70% confluence in 100 mm^3 dishes. siRNA was prepared by incubating serum free media with DharmaFECT 1 (784 µl media to 16 µl of DharmaFECT1) for 5 minutes. Gene specific siRNA was also incubated with serum free media (20 nmol siRNA in 784 µl serum free media) for 5 minutes. After

5 minutes the media/DharmaFECT mixture was added to the siRNA and allowed to incubate for 15 minutes with occasional gentle mixing. Cells to be transfected were washed with PBS and 6.4 mls of serum free media was added to the 100 mm^3 dish. The 1.6 mls of siRNA/Dharmafect1 mixture was then added to the cells and allowed to sit for 36 hours.

All cell lines used in the siRNA knockdown experiments were grown as previously described in section **2.2.1** and according to ATCC specifications (http://www.atcc.org). Cells were transfected with Dharmacon siRNA dilution buffer (mock-transfection), 20 nmol of kinase specific siRNA constructs, or with scrambled siRNA as a control. 36 hours after transfection, cells were replated in 96 well plates at a density of 2000 cells per well. RNA and protein were also harvested (as described previously) on days two and four to confirm sufficient knockdown of gene expression by Q-RT-PCR and western blotting, respectively. Remaining cells were spun down, split in two equal tubes, and RNA and protein were isolated for 0 hour knockdown readings. After replating in 96 well plates, growth was measured by MTS assay every 2 days for a total of 5 days.

2.2.5 Cell proliferation assays:

Cell growth was measured using the CellTiter 96^{TM} Aqueous Non-Radioactive Cell Proliferation assay (MTS assay, Promega) according to the manufacturer's instructions. Briefly, cells were plated in 96-well plates at 2000 cells per well. Every 24 hours, a

solution containing 20:1 ratio of MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) and PMS (phenazine methosulfate) was added to the cells. Plates were incubated at 37°C for 2 hours and absorption at 550nm was determined. Each data point was performed in heptuplicate, and the results were reported as average absorption ± standard deviation. The data is reported as percentage of growth compared to mock transfected controls for each cell line. Experiments were repeated at least twice and the percentage growth inhibition is the average of the experiments.

2.2.6. Human Breast Tumors from Baylor College of Medicine:

Human breast tumor samples were isolated from the Asterand tumor bank purchased by the Breast Center at Baylor College of Medicine. There is no clinical follow-up data for patients from whom these tumors were isolated. Tumor tissue was flash frozen and stored in liquid nitrogen until DNA, RNA, and protein isolation was performed. Tumor tissue was disrupted from the frozen human tumor samples by homogenization using a PRO Scientific rotor-stator homogenizer (Pro Scientific, Oxford, CT) with Multi-Gen7 generators. DNA was isolated from tumor cores using Qiagen's DNeasy Blood and Tissue Kit (Valencia, CA) according to the manufacturer's instructions. RNA was isolated using Qiagen's RNeasy Blood and Tissue Kit (Valencia, CA) according to the manufacturer's instructions as previously described. Protein was isolated after homogenization in protein lysis buffer consisting of 50mM HEPES pH7.5, 150mM NaCl,

1mM EDTA, 1% Triton X-100, 10% glycerol, 100mM NaF, Complete Mini protease inhibitors cocktail tablet (Roche), and phosphatase inhibitor cocktail I and II (Sigma-Aldrich). Protein concentration was determined using BCA Protein Assay Reagents (Pierce Biotechnology, Rockford, IL).

2.2.7 Human Breast Tumor Data Sets from Other Institutions:

Many of the analyses in this thesis utilized publically available datasets for analysis of gene expression, metastasis-free and overall survival, and response to therapy. A brief description of all of the data sets used in this thesis is detailed below.

Wang dataset- This dataset comes from a study entitled, "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer." [1]. The tumor bank came from the Erasmus Medical Center (Rotterdam, Netherlands) and was comprised of frozen tumor samples from patients with lymph-node-negative breast cancer who were treated during 1980–95, but who did not receive systemic neoadjuvant or adjuvant therapy. Tumors were selected so as to avoid selection bias by assuming a relapse rate of 25–30% in 5 years, and a substantial loss of tumors for quality-control reasons. Thus, 436 samples of invasive tumors were processed. Patients with poor, intermediate, and good clinical outcome were included. Samples were rejected on the basis of insufficient tumor content (53 tumors), poor RNA quality (77 tumors), or poor chip quality (20 tumors); thus, 286 samples were eligible for further analysis. The median age of the patients at surgery was 52 years (range 26–83). 219

had undergone breast-conserving surgery and 67 modified radical mastectomy. Radiotherapy was given to 248 patients (87%) according to institutional protocol. The proportions of patients who underwent breast-conserving therapy and radiotherapy are normal for lymph-node-negative disease. Patients were included irrespective of radiotherapy status because in their study, they did not aim to investigate the effects of a specific type of surgery or adjuvant radiotherapy. Lymph-node negativity was based on pathological examination by regional pathologists. All 286 tumor samples were confirmed to have sufficient (>70%) tumor and uniform involvement of tumor in 5 µm frozen sections stained with hematoxylin and eosin (H&E). Amounts of estrogen receptor alpha (hereafter referred to simply as ER) and progesterone receptors (PR) were measured by ligand-binding assay, EIA, or immunohistochemistry (nine tumors). The cut-off value for classification of patients as positive or negative for ER and PR was 10 fmol per mg protein or 10% positive tumor cells. Postoperative follow-up involved examinations every 3 months for 2 years, every 6 months for years 3–5, and every 12 months from year 5. The date of diagnosis of metastasis was defined as that at confirmation of metastasis after symptoms reported by the patient, detection of clinical signs, or at regular follow-up.

van de Vijver dataset- This dataset came from a study entitled, "A gene-expression signature as a predictor of survival in breast cancer" by van de Vijver *et al*. [2]. Tumors from a series of 295 consecutive women with breast cancer were selected from the fresh frozen tissue bank of the Netherlands Cancer Institute according to the following

criteria: the tumor was primary invasive breast carcinoma that was less than 5 cm in diameter at pathological examination (pT1 or pT2); the apical axillary lymph nodes were tumor-negative, as determined by a biopsy of the infraclavicular lymph nodes; the age at diagnosis was 52 years or younger; the calendar year of diagnosis was between 1984 and 1995; and there was no previous history of cancer, except non-melanoma skin cancer. All patients had been treated by modified radical mastectomy or breast-conserving surgery, including dissection of the axillary lymph nodes, followed by radiotherapy if indicated. Among the 295 patients, 151 had lymph node–negative disease (results on pathological examination, pN0) and 144 had lymph-node positive disease (pN+). Ten of the 151 patients who had lymph-node–negative disease and 120 of the 144 who had lymph-node–positive disease had received adjuvant systemic therapy consisting of chemotherapy (90 patients), hormonal therapy (20 patients), or both (20 patients). All patients were assessed at least annually for a period of at least five years. The median duration of follow-up was 7.8 years (range, 0.05 to 18.3) for the 207 patients without metastasis as the first event and 2.7 years (range, 0.3 to 14.0) for the 88 patients with metastasis as the first event. The median follow-up among all 295 patients was 6.7 years (range, 0.05 to 18.3). There were no missing data. The level of expression of estrogen receptor alpha was estimated on the basis of the hybridization results on the microarray experiments, which is a reliable assay for estrogen-receptor status. On the basis of this assay, there were 69 ER–negative tumors (defined by an intensity ratio of less than –0.65 U on a logarithmic scale, corresponding to staining of

less than 10 percent of nuclei on immunohistochemical analysis) and 226 ER–positive tumors in the cohort.

Desmedt dataset- This dataset came from a study entitled, "Strong Time Dependence of the 76-Gene Prognostic Signature for Node-Negative Breast Cancer Patients in the TRANSBIG Multicenter Independent Validation Series." by Desmedt *et al*. [3]. Patients were eligible for inclusion if they were younger than 61 years old at diagnosis, diagnosed before 1999 with node-negative, T1–T2 (5 cm) breast cancer, and had not received adjuvant systemic therapy. Patients with previous malignancies (except basal cell carcinoma) or with bilateral synchronous breast tumors were excluded. A total of 326 patients were included. Patients in this series had been diagnosed between 1980 and 1998 and had a median follow-up of 13.6 years. Data were also available for the 151 node-negative patients included in the analyses carried out at the Netherlands Cancer Institute (NKI). Initially, frozen samples from eligible patients (n = 403) were sent from all clinical centers to NKI for RNA extraction and microarray analysis. Useful RNA could be extracted for hybridization and analysis from 81% of these frozen samples, leaving 326 samples available for analysis. Paraffin-embedded tumor samples from all patients in both the original and the validation series were sent to the Department of Pathology at the European Institute of Oncology, Milan, where the same pathologist (GV) determined ER status [using immunohistochemistry] and histologic grade [using the Elston and Ellis method]. Whenever possible, these central pathology data were

used to determine ER status (n = 218) and histologic grade (n = 237); otherwise the local pathology data obtained at the original clinical center were used.

Ivshina dataset- This dataset came from a paper entitled, "Genetic Reclassification of Histologic Grade Delineates New Clinical Subtypes of Breast Cancer" published by Ivshina *et al*. [4]. This tumor set included three cohorts of patients.

Cohort 1-Uppsala cohort: The Uppsala cohort originally composed of 315 women representing 65% of all breast cancers resected in Uppsala County, Sweden, from January 1, 1987, to December 31, 1989. For histologic grading, new tumor sections were prepared from the original paraffin blocks and stained with eosin (with the exception of a few original van Gieson-stained sections). All sections were graded in a blinded fashion (H.N.) according to the Nottingham Grading System as follows: Tubule Formation: 3 = poor, if <10% of the tumor showed definite tubule formation, 2 = moderate, if 10% but 75%, and 1 = well, if >75%. Mitotic Index: 1 = low, if <10 mitoses, 2 = medium, if 10 to 18 mitoses, and 3 = high, if >18 mitoses (per 10 high-power fields). Field diameter was 0.57 mm. Nuclear Grade: 1 = low, for little variation in size and shape of nuclei, 2 = medium for moderate variation, and 3 = high for marked variation and large size. Tumors with summed scores ranging from 3 to 5 were classified as G1; 6 to 7 as G2; and 8 to 9 as G3. Estrogen and progesterone receptors were assessed by Abbott's quantitative enzyme immunoassay (Abbott Laboratories, Chicago, IL) and deemed positive if >0.05 fmol/μg DNA. Vascular endothelial growth factor (VEGF) was measured in tumor cytosol by a quantitative immunoassay kit (Quantikine-human VEGF;

R&D Systems, Minneapolis, MN). Protein levels of Ki67 were analyzed using anti-Ki67 antibody (MIB-1) by the grid-graticula method with cutoffs: low = 2, medium >2 and <6, high = 6. Cyclin E was measured using the antibody HE12 (Santa Cruz Biotechnology, Inc., Santa Cruz, CA) with cutoffs: low = 0% to 4%, medium = 5% to 49%, and high = 50% to 100% stained tumor cells (22). Vascular growth was determined by routing staining of tumor sections. P53 mutational status was determined by cDNA sequencing.

Cohort 2-Stockholm cohort: The Stockholm samples were derived from breast cancer patients operated on at the Karolinska Hospital from January 1, 1994, through December 31, 1996, and identified in the Stockholm-Gotland breast cancer registry. Information on patient age, tumor size, number of metastatic axillary lymph nodes, hormonal receptor status, distant metastases, site and date of relapse, initial therapy, and date and cause of death were obtained from patient records and the Stockholm-Gotland Breast Cancer Registry. Tumor sections were graded in the same fashion as the Uppsala tumors. Only histologic G2 samples were evaluated in this study.

Cohort 3-Singapore cohort: The Singapore samples were derived from patients operated on at the National University Hospital (Singapore) from February 1, 2000, through January 31, 2002. Routine clinical data were obtained from pathology reports, but no information on recurrence or cause of death was available. Tumor sections were graded according to the Nottingham grading system as applied to the Uppsala and Stockholm cohorts, with the following exception: Mitotic Index: 1 = low, if <8 mitoses; 2 = medium, if 9 to 16 mitoses; and 3 = high, if >16 mitoses (per 10 high-power fields);

field diameter was 0.55 mm. Only histologic G2 samples were evaluated in this study. For complete information of these datasets refer to **Figure 2.1.**

Denmark (training) dataset- The training dataset used for much of the proteomic studies are detailed in the Specific Material and Methods of **Chapter 4.**

Superslide (validation) dataset- The Superslide dataset used for much of the validation of the proteomic studies are detailed in the Specific Material and Methods of **Chapter 4**.

Figure 2.1

	Uppsala n=249			Stockholm n=58	Singapore n=40
Variables, by grade	**G1** n=68	**G2** n=126	**G3** n=55	**G2** n=58	**G2** n=40
Age, median yrs	62	63	62	58	52
<55 years, %	26	25	44	41	60
Tumor size, cm	1.8	2.2	2.9	2.5	2.8
Nodes, positive, %	15	35	55	50	40
ER negative tumors, %	3	9	38	7	28
Follow up, median yrs	11	9	6	7	-
All recurrences, %	26	39	50	24	-
Endocrine therapy, %	18	37	36	62	-
Chemotherapy, %	4	6	22	5	-
Combine therapy, %	2	3	0	16	-
No systemic therapy, %	77	54	45.5	17	

Figure 2.1 Clinical Characteristics of Tumors included in the Ivshina *et al.* Dataset- The clinical characteristics of the tumors included in the Ivshina dataset used for analysis of MELK expression on overall survival.

2.3 Specific Materials and Methods

2.3.1 The Methods of Chapter 3:

2.3.2 Study population and design:

A total of 102 patients with invasive breast cancer were recruited through IRB-approved, neoadjuvant studies to investigate gene expression in human tumors before and after drug treatment. These tumors were collected by Dr. Jenny Chang at Baylor College of Medicine and were graciously provided for further analysis. Breast biopsies using a core needle were taken before initiation of any treatment and were used in this study. Because the patients did not receive systemic adjuvant or neoadjuvant therapy prior to the biopsy, the results from the gene expression analysis represent basal gene expression in these breast cancers. For these gene expression profiling experiments, 102 breast tumors were studied, 58 of which were ER-positive and 44 ER-negative by IHC-staining (24 of which were confirmed as “triple-negative”). The tumors were all stage III or IV from pre- and post-menopausal women, with all tumors showing >30% cellularity. The women were from several racial groups and the majority had no palpable nodes at baseline. Most of the women were premenopausal and presented with relatively large tumors (ranging from 2.5 to 25 cm). The clinical and demographic features of these tumors are summarized in **Table 2.2**.

Table 2.2. Clinical characteristics of the patients and tumor samples used in the study.

Characteristic	Tumor Set N=102 (%)
Age	
Mean	48.1
Range	(32-72)
Race	
Caucasian	50 (57%)
Hispanic	7 (8%)
African-American	23 (27%)
Asian	7 (8%)
Menopausal Status	
Pre	49 (62%)
Post	30 (38%)
BMI	
Mean	29.7
Range	(16.1-48.3)
Baseline Tumor Size, cm	
Mean	6.3
Range	(2.5-25.0)
Palpable Nodes at Baseline	
Yes	20 (21%)
No	77 (79%)
ER	
Positive	57 (56%)
Negative	45 (44%)
Unknown	0 (0%)
PR	
Positive	37 (36%)
Negative	47 (46%)
Unknown	18 (18%)
HER2/Neu	
Positive	27 (26%)
Negative	58 (57%)
Unknown	17 (17%)

Table 2.2. Characteristics of 102 patients with breast cancer. Tumors from these patients were collected by Dr. Jenny Chang at Baylor College of Medicine and used for gene expression profiling to identify overexpressed kinases in ER-negative breast tumors

2.3.3 Acquisition of breast tumor tissue:

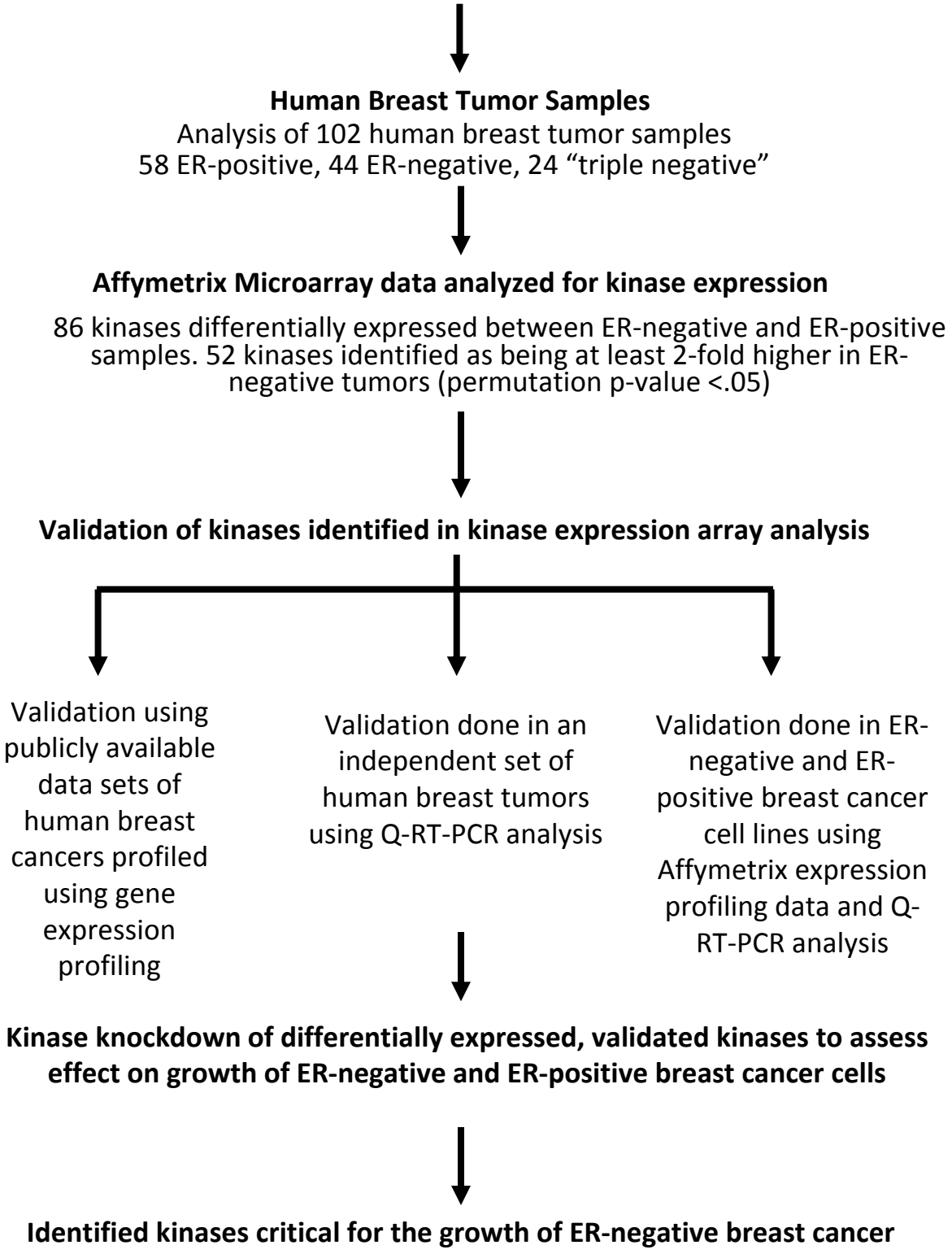
All ER-negative and ER-positive tumors were collected by Dr. Jenny Chang at Baylor College of Medicine through IRB-approved, neoadjuvant studies to investigate gene expression changes in human tumors following drug treatment. Diagnostic core needle biopsies were taken first, then several (up to 6) additional cores were taken for biomarker studies. These additional cores were taken before treatment, placed immediately in liquid nitrogen, and used to prepare RNA, DNA, and protein. Immunohistochemical (IHC) staining for ER alpha and *HER2/neu* expression was done on these sets of tumor samples as previously described [5]. The tumor set comprised of pre-treatment specimens from studies of docetaxel [6], cyclophosphamide [7], docetaxel, and cyclophosphamide (unpublished data), and trastuzumab [8]. All studies were conducted with approval from the Institutional Review Boards at Baylor College of Medicine and participating sites. An outline of the study design is depicted in **Figure 2.2**.

2.3.4 Affymetrix microarray experiments:

All Affymetrix gene expression studies were conducted in the laboratory of Dr. Jenny Chang at Baylor College of Medicine and data was graciously provided by Dr. Jenny Chang. Total RNA from tumor samples was isolated using Qiagen's RNeasy kit, double-stranded cDNA synthesized, and reverse transcription carried out followed by biotin labeling. Additionally, about 250-fold linear amplification and phenol-chloroform

Figure 2.2 Outline of the study design used in the kinome expression profiling in breast cancer. 102 human breast tumors were collected and gene expression profiling was performed to identify kinases overexpressed in ER-negative breast cancer by Dr. Jenny Chang at Baylor College of Medicine. These overexpressed kinases were confirmed two separate ways using: 1)Q-RT-PCR and 2) gene expression profiling on human breast tumors and breast cancel cell lines. siRNA knockdown of validated, overexpressed kinases identified those kinases that were critical for mitogenesis in ER-negative breast cancer. These kinases are tractable targets for the treatment of ER-negative breast cancer.

Figure 2.2 Human Breast Tumor Acquisition



cleanup was done as previously published [9]. From each biopsy, 15 micrograms of biotin-labeled cRNA was hybridized onto an Affymetrix HGU133A GeneChip^{TM}, which comprise around 22,000 genes (www.affymetrix.com). The experiments were all done using the microarray core facility at the Lester and Sue Smith Breast Center at Baylor College of Medicine.

2.3.5 Statistical analysis of microarray data:

Statistical analysis was done with dChip (www.dchip.org) and BRB ArrayTools software packages developed by Dr. Richard Simon and Amy Peng Lam. (http://linus.nci.nih.gov/BRB-ArrayTools.html). Subsequent data analysis and clustering was limited to the known kinome with interrogation of the 779 known and putative human protein, nucleotide, and lipid kinases as well as kinase-interacting proteins and regulatory subunits as previously described [10-12]. Gene expression was estimated with dChip software using Invariant Set normalization and Perfect Match (PM) only model [13]. Comparison of ER-negative and ER-positive groups was done with BRB Array Tools, using t-test and computing permutation p-values [14]. Hierarchical clustering was also done using dChip and MeV software packages (http://www.tm4.org/mev.html) with rows standardized by subtracting the mean and dividing by the standard deviation. Pearson's correlation and centroid linkage was used to generate the trees on Log2 transformed expression data with perfect match/mismatch (PM/MM) difference background subtraction. Original data analysis in dChip and with BRB array tools was

done by Dr. Anna Tsimelzon and Dr. Susan Hilsenbeck. Reanalysis with MeV and dChip and further bioinformatic analysis was done by Corey Speers.

2.3.6 Gene ontology analysis:

All gene ontology enrichment analyses were initially done using a Pathway Architect™ software package developed by Stratagene by Corey Speers. Genes found to be overexpressed at least 2-fold with a permutation p-value score of <.05 were used as the input list and compared against the human kinome. Follow-up and confirmatory analysis was done using Gene Ontology Tree Maker (GOTM) and EASE software [15].

2.3.7 Selection of genes for further study:

After completing all microarray experiments and performing statistical analysis including the generation of gene cluster groups, candidate genes were validated in studies as described below. Selected candidate genes had a > 2.0 fold increase in ER-negative vs. ER-positive tumors with a p value <.05 and a false discovery rate (FDR) of 1%. Additional statistical analysis were done to minimize false discoveries and limit misclassification as described by Dalmasso [16, 17]. The aberrant expression of these selected genes was then confirmed using Q-RT-PCR on the original set of RNA from the human breast tumor samples, and was then analyzed in our validation studies. We then used comprehensive bioinformatics analysis to choose promising targets. For all of our high throughput analysis, we employed all appropriate quality control and normalization

techniques (QC-RMA, MAS5) before proceeding to data exploration. This exploration involved unsupervised data analysis using hierarchical and k-means clustering, principal component analysis and ordination with Pearson correlation coefficient, Spearman rank, as well as Eigen plot analysis. Dimension reduction ordination with Eigen plots allowed us to investigate the data in more meaningful ways by discovering hidden associations. Integration of datasets using coinertia analysis allowed us to integrate data across platforms and this process, coupled with GO ontology analysis and categorization, lead to additional target discovery. In effect this allowed us to go beyond a statistical means of target selection and allowed us to integrate multiple data in picking targets that are common between all datasets. These analyses were done by Corey Speers with assistance of Dr. Susan Hilsenbeck at Baylor College of Medicine and Dr. John Quackenbush at the Dana Farber Cancer Institute.

2.3.9 Z-transform test in multiple datasets:

To validate that the differentially expressed kinases identified in this analysis were also differentially expressed in other publically available datasets of human breast cancer we employed the Z-transform test described by Whitlock [18]. Briefly, this method allows for the combining of individual *P*-values and has proven superior to Fisher's combined probability test. The *Z*-transform test takes advantage of the one-to-one mapping of the standard normal curve to the *P*-value of a one-tailed test. As *Z* goes from negative infinity to infinity, *P* will go from 0 to 1, and any value of *P* will uniquely

be matched with a value of *Z* and *vice versa*. The *Z*-transform test converts the one-tailed *P*-values, P_i, from each of *k* independent tests into standard normal deviates Z_i. The sum of these Z_i's, divided by the square root of the number of tests, *k*, has a standard normal distribution if the common null hypothesis is true. The equation: $Z_S = \frac{\sum_{i=1}^{k} Z_i}{\sqrt{k}}$ was used in the calculation of summed z-scores, which were then related to the reported *P*-values. This analysis was done by Corey Speers with input from Dr. Susan Hilsenbeck at Baylor College of Medicine.

2.3.10 Kaplan-Meier Survival Analysis:

Gene expression profiling and survival data generated by Wang *et al*. and van de Vijver *et al.* was used to evaluate prognostic import of the kinase clusters in this data set [1, 2]. Data was obtained from GEO and hierarchical clustering performed only on the ER-negative samples from the Wang and van de Vijver data sets. MeV and R software package were used for data and statistical analysis. For hierarchical clustering, the expression values of the kinases identified as being over expressed in ER-negative tumors were extracted from the data sets using Affymetrix probe IDs. The expression values were mean centered and hierarchical clustering based on Pearson's correlation with complete linkage again identified our 4 subsets of ER-negative kinase clusters. Figure of merit scoring showed that these four clusters were stable against reclustering in both datasets. Using this information each tumor sample was classified as falling into

one of the 4 kinase clusters (cell cycle checkpoint, S6 kinase, MAPK signaling, or immunomodulatory). After classification of tumors, Kaplan-Meier analysis using the survival data from the ER-negative tumors in the data sets was performed using R (http://www.r-project.org) and survival curves were generated. Chi squared scores were calculated to determine significance. These analyses were done by Corey Speers with assistance of Dr. Susan Hilsenbeck at Baylor College of Medicine.

2.4 The Materials and Methods of Chapter 4.

2.4.1 Training and Validation Datasets used in the Proteomic Analysis:

Two datasets (a training and a validation set) were used in this analysis. These tumors were collected by investigators in Denmark and at M.D. Anderson Cancer Center under the direction of Dr. Gordon Mills, and the data was generously provided to us for further analysis. The training set was derived from tumors acquired in collaboration with investigators in Denmark, and is at times referred to as the Denmark dataset. This dataset is comprised of 166 tumors. The mean age of the women from whom the tumors were derived was 54.8 (range 30-69). 126 of the tumors were ER-positive (76%) and 40 tumors were ER-negative (24%). 34 (21%) tumors were HER2-positive as measured by fluorescence *in situ* hybridization (FISH) analysis and 102 (61%) were HER2-negative, with 30 (18%) having unknown HER2 status. The majority of the women (93%) from whom these tumors were derived had node-positive disease, with one to three positive nodes in 46% of women and greater than 3 nodes positive in 47% of women. The mean survival of these women was 107.5 months. Please refer to **Table 4.1** in **Chapter 4** for more details of this tumor set.

The validation dataset was comprised of 712 tumors (612 tumors from M.D. Anderson Cancer Center and 91 tumors from Baylor College of Medicine). The tumor samples obtained and processed at M.D. Anderson by Dr. Gordon Mills, and had clinical follow-up data including date of diagnosis, disease and recurrence free survival,

metastasis-free survival, and overall survival. The tumor samples obtained at Baylor College of Medicine were processed by Corey Speers and had initial diagnosis data but did not have clinical follow-up data and thus were censored from any analysis involving clinical follow-up data. The mean age of the women from whom the tumors were derived was 60.3 (range 23-89). 449 of the tumors were ER-positive (63%) and 263 tumors were ER-negative (37%). 21 (3%) tumors were HER2-positive as measured by fluorescence *in situ* hybridization (FISH) analysis and 148 (21%) were HER2-negative, with 543 (76%) having unknown HER2 status. 38% of the women had node-negative disease, while 28% had 1-3 nodes positive and 11% had greater than 3 nodes positive. The mean survival of these women was 69.86 months. Please refer to **Table 4.1** in **Chapter 4** for additional details of this tumor set.

2.4.2 Reverse phase protein lysate arrays (RPPAs):

This assay is a high throughput proteomic technique to measure protein expression of hundreds of proteins in thousands of samples simultaneously. These experiments were done in the laboratory of Dr. Gordon Mills using his proteomic facilities. Dr. Bryan Hennessy did the proteomic analysis at M.D. Anderson and kindly provided the data for our analysis. Briefly, lysis buffer was used to lyse cell lines and frozen human tumor samples by homogenization. Cell and tumor lysates were normalized to 1 μg/μl concentration using bicinchronic acid assay and boiled with 1% SDS, and the supernatants were manually diluted in six or eight 2-fold dilutions with lysis

buffer in Dr. Mills' facility. A GeneTAC arrayer (Genomic Solutions, Inc.) created 1,152 spot arrays on nitrocellulose-coated FAST slides (Sleicher & Schuell Biosciences, Inc.) from the serial dilutions. Up to 7000 single dots were printed on one slide allowing for the analysis of up to 1054 samples with 1000 controls on a single slide (a lower density approach is presented in **Figure 2.3** for visualization). The serial dilution provides a slope and intercept, allowing relative quantification of individual proteins (**Figure 2.3**). This is compared with control peptides for total and phosphopeptides allowing absolute quantification. With robotics, 100 identical slides can be printed at one time. Each slide was probed with a validated antibody (listed in **Table 2.3**). The DAKO signal amplification system was used to detect and amplify AB-binding intensity. This is a commercially available catalyzed system kit that uses 3,3'-diaminobenzidine tetrachloride and a catalyzed reporter deposition of the substrate to amplify the signal detected by the primary antibody. A biotinylated secondary antibody (anti-mouse or anti–rabbit) is used as a starting point for signal amplification. A streptavidin-biotin complex attached to the secondary antibody and biotinyl-tyramide deposition on this complex was used to amplify the reaction. Tyramide-bound horseradish peroxidase cleaves 3,3'-diaminobenzidine tetrachloride, resulting in a stable brown precipitate with excellent signal-to-noise ratio. This technique is sensitive and reproducible in the femtomolar sensitivity range.

The slides were scanned, analyzed, and quantitated using Microvigene software (VigeneTech Inc.) to generate serial dilution–signal intensity curves for each sample with

Figure 2.3

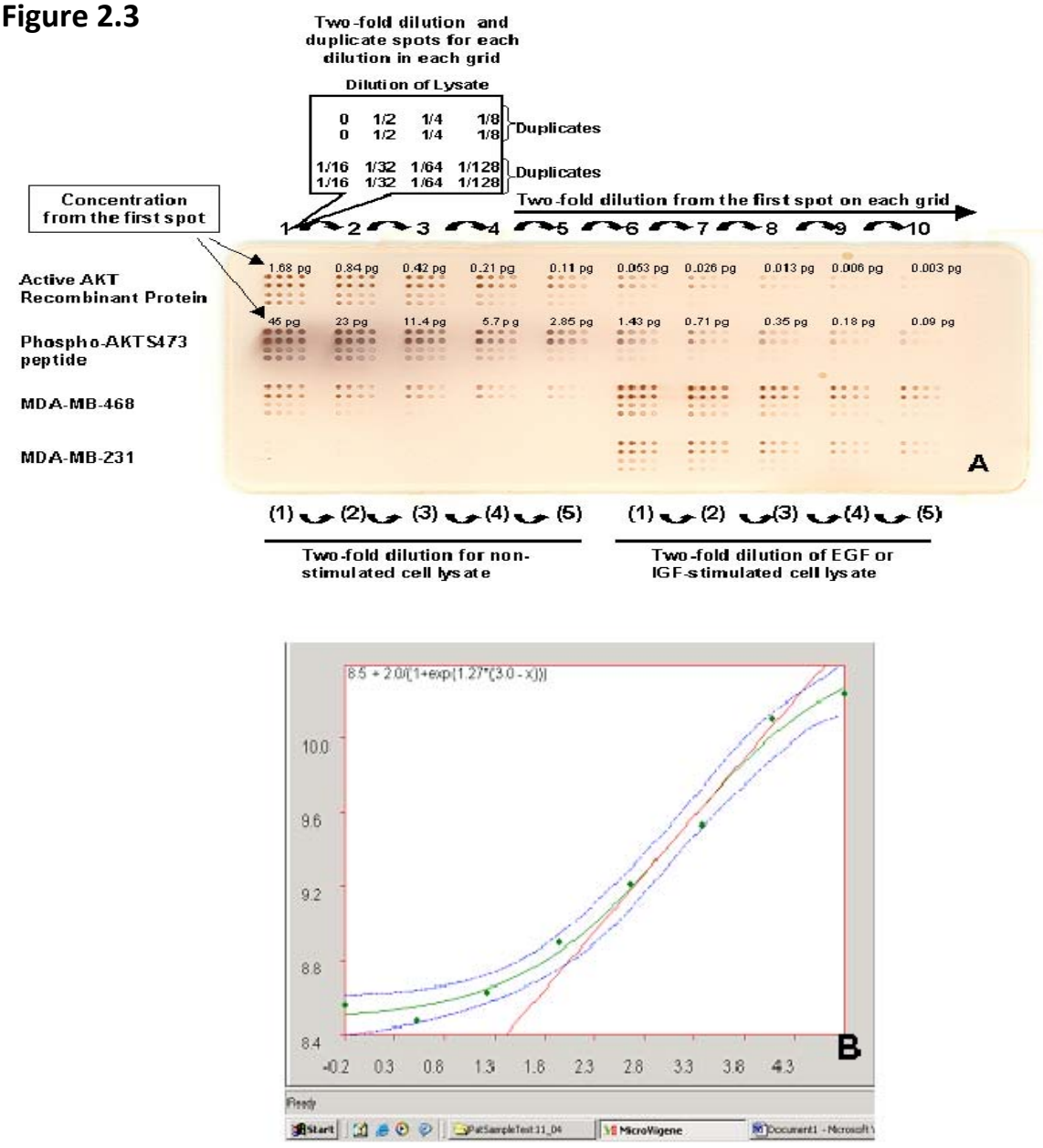


Figure 2.3 Reverse phase protein lysate array. Samples are duplicate spotted in 6-8 serial dilutions on nitrocellulose slides **(A).** Slides are scanned and Microvigene software generates concentration curves for individual samples using 'dot' intensity **(B). Slide used with permission by Dr. Bryan Hennessy**

Table 2.3. Eighty-nine monospecific antibodies used in this study.

Antibody name	Protein name	Company	cat#	Host	Dilution
4EBP1	4E Binding Protein 1	Cell Signaling Technology, Inc.	CS 9452	Rabbit	1 in 100
4EBP1p37	4EBP1 phosphorylation at T37/T46	Cell Signaling Technology, Inc.	CS 9459	Rabbit	1 in 100
AcCoA	Acetyl CoA Carboxylase	Epitomics, Inc.	1768-1	Rabbit	1 in 250
AcCoAp	AcCoA phosphorylation at S79	Cell Signaling Technology, Inc.	CS 3661	Rabbit	1 in 250
Akt	Protein Kinase B	Cell Signaling Technology, Inc.	CS 9272	Rabbit	1 in 250
Aktp308	Akt phosphorylation at S308	Cell Signaling Technology, Inc.	CS 9275	Rabbit	1 in 250
Aktp473	Akt phosphorylation at S473	Cell Signaling Technology, Inc.	CS 9271	Rabbit	1 in 250
AMPK	AMPK	Cell Signaling Technology, Inc.	CS 2532	Rabbit	1 in 250
AMPKp	AMPK phosphorylation at S172	Cell Signaling Technology, Inc.	CS 2535	Rabbit	1 in 250
AR	Androgen receptor	Epitomics	1852-1	Rabbit	1 in 200
β catenin	B catenin	Cell Signaling Technology, Inc.	CS 9562	Rabbit	1 in 300
bcl2	bcl2	Dako	M0887	Mouse	1 in 200
BRCA1	BRCA1	Upstate Biotechnology, Inc.	07-434	Rabbit	1 in 1000
caveolin 1	Caveolin 1	Cell Signaling Technology, Inc.	CS 3232	Rabbit	1 in 250
CCNB1	Cyclin B1	Epitomics, Inc.	1495-1	Rabbit	1 in 500
CCND1	Cyclin D1	Santa Cruz Biotechnology, Inc.	SC-718	Rabbit	1 in 1000
CCNE1	Cyclin E1	Santa Cruz Biotechnology, Inc.	SC-247	Mouse	1 in 500
CCNE2	Cyclin E2	Epitomics	1142-1	Rabbit	1 in 250
CD31	CD31	Dako	M0823	Mouse	1 in 500
CDK4	CDK4	Cell Signaling Technology, Inc.	CS 2906	Rabbit	1 in 250
cjun	Cjun	Cell Signaling Technology, Inc.	CS 9165	Rabbit	1 in 250
ckit	Ckit	Cell Signaling Technology, Inc.		Rabbit	1 in 150
cleaved caspase 7	Cleaved caspase 7 (Asp198)	Cell Signaling Technology, Inc.	CS 9491	Rabbit	1 in 150
cleaved PARP	Cleaved PARP (Asp214)	Cell Signaling Technology, Inc.	CS 9546	Mouse	1 in 250
cmyc	Cmyc	Cell Signaling Technology, Inc.	CS 9402	Rabbit	1 in 150
Collagen VI	Collagen VI	Santa Cruz Biotechnology, Inc.	SC-20649	Rabbit	1 in 750
COX2	COX2	Epitomics, Inc.	2169-1	Rabbit	1 in 500
E cadherin	E cadherin	Cell Signaling Technology, Inc.	CS 4065	Rabbit	1 in 200
EGFR	Epidermal growth factor receptor	Santa Cruz Biotechnology, Inc.	SC-03	Rabbit	1 in 200
EGFRp1045	EGFR phosphorylation at Y1045	Cell Signaling Technology, Inc.	CS 2237	Rabbit	1 in 100

EGFRp922	EGFR phosphorylation at Y992	Cell Signaling Technology, Inc.	CS 2235	Rabbit	1 in 100
ER	Estrogen receptor alpha	Lab Vision Coorporation (formerly Neomarkers)	Sp1	Rabbit	1 in 250
ERK2	Mitogen-activated protein kinase	Cell Signaling Technology, Inc.	SC-154	Rabbit	1 in 250
ERp118	ER phosphorylation at S118	Epitomics, Inc.	1091-1	Rabbit	1 in 200
ERp167	ER phosphorylation at S167	Epitomics, Inc.	2492-1	Rabbit	1 in 200
FGFR1	FGFR1	Santa Cruz	SC-7945	Rabbit	1 in 250
GATA3	GATA binding protein 3	BD Biosciences	558686	Mouse	1 in 200
GSK3	Glycogen synthase kinase 3 beta	Santa Cruz Biotechnology, Inc.	SC-7291	Mouse	1 in 1000
GSK3p21_9	GSK3 phosphorylation at S21/S9	Cell Signaling Technology, Inc.	CS 9331	Rabbit	1 in 250
HER2	Human epidermal receptor 2	Epitomics, Inc.	1148-1	Rabbit	1 in 250
HER2p1248	HER2 phosphorylation at Y1248	Upstate Biotechnology, Inc.	06-229	Rabbit	1 in 750
IGF1R	Insulin-like growth factor receptor 1	Cell Signaling Technology, Inc.	CS 3027	Rabbit	1 in 500
IGFRp	IGF1R phosphorylation at Y1135/Y1136	Cell Signaling Technology, Inc.	CS 3024	Rabbit	1 in 200
JNK	cjun N terminal Kinase	Santa Cruz Biotechnology, Inc.	SC-474	Rabbit	1 in 200
JNKp183-185	JNK phosphorylation at T183/Y185	Cell Signaling Technology, Inc.	CS 9251	Rabbit	1 in 150
LKB1	LKB1	Abcam	15095	Mouse	1 in 200
LKB1p	LKB1p	Cell Signaling Technology, Inc	SC-3054	Rabbit	1 in 250
MAPKp	MAPK1/2 phosphorylation at T202/T204	Cell Signaling Technology, Inc.	CS 4377	Rabbit	1 in 1000
MEK1	MAPK/ERK kinase 1	Epitomics, Inc.	1235-1	Rabbit	1 in 15000
MEK12p	MEK1/2 phosphorylation at T217/T221	Cell Signaling Technology, Inc.	CS 9121	Rabbit	1 in 800
mTOR	mammalian target of rapamycin	Cell Signaling Technology, Inc.	CS 2983	Rabbit	1 in 400
p110alpha	p110alpha subunit of phosphatidylinositol-3-kinase	Epitomics, Inc.	1683-1	Rabbit	1 in 500
p21	p21	Santa Cruz Biotechnology, Inc.	SC-397	Rabbit	1 in 250
p27	p27	Santa Cruz Biotechnology, Inc.	SC-527	Rabbit	1 in 500
p38	p38 MAPK	Cell Signaling Technology, Inc.	CS 9212	Rabbit	1 in 300
p38p180_2	p38 MAPK phosphorylation at T180/T182	Cell Signaling Technology, Inc.	CS 9211	Rabbit	1 in 250
p53	p53	Cell Signaling Technology, Inc.	CS 9282	Rabbit	1 in 3000
p7056 Kinase	p70S6 Kinase	Epitomics, Inc.	1494-1	Rabbit	1 in 500
p70S6Kp389	p70S6 Kinase phosphorylation at T389	Cell Signaling Technology, Inc.	CS 9205	Rabbit	1 in 200
PAI1	Plasminogen activator inhibitor-1	BD Biosciences	612024	Mouse	1 in 1000
pcmyc	cmyc phosphorylation at T58/S62	Cell Signaling Technology, Inc.	CS 9401	Rabbit	1 in 150

PDK1	Phosphoinositide Dependent Kinase 1	Cell Signaling Technology, Inc.	CS 3062	Rabbit	1 in 250
PDK1p241	PDK1 phosphorylation at S241	Cell Signaling Technology, Inc.	CS 3061	Rabbit	1 in 500
PKCalpha	Protein Kinase C alpha	Upstate Biotechnology, Inc.	05-154	Mouse	1 in 2000
PKCaphap657	PKCalpha phosphorylation at S657	Upstate Biotechnology, Inc.	06-822	Rabbit	1 in 3000
pmTOR	mTOR phosphorylation at S2448	Cell Signaling Technology, Inc.	CS 2971	Rabbit	1 in 150
PR	Progesterone receptor	Epitomics, Inc.	1483-1	Rabbit	1 in 400
PTEN	PTEN	Cell Signaling Technology, Inc.	CS 9552	Rabbit	1 in 500
Rab25	Rab25	Courtesy Dr. Kwai Wa Cheng, MDACC	Covance	Rabbit	1 in 4000
Rb	Retinoblastoma	Cell Signaling Technology, Inc.	CS 9309	Mouse	1 in 3000
Rbp	Rb phosphorylation at S807/S811	Cell Signaling Technology, Inc.	CS 9308	Rabbit	1 in 250
S6	S6 ribosomal protein	Cell Signaling Technology, Inc.	CS 2217	Rabbit	1 in 200
S6p235-236	S6 phosphorylation at S235/S236	Cell Signaling Technology, Inc.	CS 2211	Rabbit	1 in 3000
S6p240_4	S6 phosphorylation at S240/S244	Cell Signaling Technology, Inc.	CS 2215	Rabbit	1 in 3000
SGK	Serum Glucocorticoid Kinase	Cell Signaling Technology, Inc.	CS 3272	Rabbit	1 in 250
SGKp	SGK phosphorylation at S78	Cell Signaling Technology, Inc.	CS 3271	Rabbit	1 in 250
src	Src	Upstate Biotechnology, Inc.	05-184	Mouse	1 in 200
srcp416	src phosphorylation at Y416	Cell Signaling Technology, Inc.	CS 2101	Rabbit	1 in 150
srcp527	src phosphorylation at Y527	Cell Signaling Technology, Inc.	CS 2105	Rabbit	1 in 400
stat3	STAT3	Upstate Biotechnology, Inc.	06-596	Rabbit	1 in 500
stat3p705	stat3 phosphorylation at S705	Cell Signaling Technology, Inc.	CS 9131	Rabbit	1 in 500
stat3p727	stat3 phosphorylation at S727	Cell Signaling Technology, Inc.	CS 9134	Rabbit	1 in 250
stat6p641	stat6 phosphorylation at Y641	Cell Signaling Technology, Inc.	CS 9361	Rabbit	1 in 150
stathmin	Stathmin	Epitomics, Inc.	1972-1	Rabbit	1 in 500
TSC2	Tuberous Sclerosis Kinase 2	Epitomics, Inc.	1613-1	Rabbit	1 in 500
TSC2p	TSC2 phosphorylation at T1462	Cell Signaling Technology, Inc.	CS 3617	Rabbit	1 in 200
VEGFR2	KDR2 / VEGF Receptor 2	Cell Signaling Technology, Inc.	CS 2479	Rabbit	1 in 700
XIAP	X linked inhibitor of apoptosis	Cell Signaling Technology, Inc.	CS 2042	Rabbit	1 in 200

Companies
Abcam, Inc. (Cambridge, MA), BD Biosciences (San Jose, CA), Cell Signaling Technology, Inc. (Danvers, MA), Dako (Carpinteria, CA), Epitomics, Inc. (Burlingame, CA), Santa Cruz Biotechnology, Inc. (Santa Cruz, CA), Upstate Biotechnology, Inc. (Millipore)
All RPPA experiments were done by Dr. Bryan Hennessy in the laboratory of Dr. Gordon Mills at M.D. Anderson Cancer Center.

the logistic fit model: $\ln(y) = a + (b - a) / (1 + \exp\{c*[d - \ln(x)]\})$ as described previously [19]. A representative natural logarithmic value of each sample curve on the slide (curve average) was then used as a relative quantification of the amount of each protein in each sample. Protein loading was corrected across samples by correction of the linear expression values using the average expression levels of at least four proteins (e.g., ERK2, GSK3, JNK, mTor) to calculate a loading correction factor for each sample. To accurately determine the absolute concentrations of proteins in a sample, standard signal intensity-concentration curves for purified proteins or recombinant peptides of known concentration was generated for comparison with the samples in which protein concentrations are unknown. Determination of the protein concentration in control cell lysates (e.g., MDA-MB-468) was very important, as it served as a reference curve on each slide for inter-slide comparison of the samples. As mentioned earlier, this work was done by Dr. Bryan Hennessy in Dr. Gordon Mills' laboratory at M.D. Anderson Cancer Center.

2.4.3 Western blot analysis:

For protein isolation from tissue culture cell lines, cells were washed once with ice-cold phosphate buffered saline (PBS) and lysed in protein lysis buffer consisting of 50mM HEPES pH7.5, 150mM NaCl, 1mM EDTA, 1% Triton X-100, 10% glycerol, 100mM NaF, Complete Mini protease inhibitors cocktail tablet (Roche), and phosphatase inhibitor cocktail I and II (Sigma-Aldrich). Protein concentration was determined using

BCA Protein Assay Reagents (Pierce Biotechnology). Western blot analysis was performed as previously described [20]. Briefly, aliquots of total protein (30μg) were resolved by electrophoresis in 10% SDS-PAGE gel and transferred to a nitrocellulose membrane (Amersham Biosciences). The membrane was blocked and incubated with primary antibody. After washing in TBST, the membrane was then incubated with horseradish peroxidase-conjugated secondary, washed again, and antigen-antibody complexes were detected using the ECL or ECL Plus chemiluminescent system (Amersham Bioscience). Primary antibodies specific for kinases of interest were purchased from Cell Signaling or Novacastra. Anti-mouse and anti-rabbit secondary antibodies were obtained from Amersham Bioscience. Protein isolation from human tumor samples was identical to that of cells, except that samples were first homogenized using a 7 mm generator and a rotator-stator homogenizer (ProScientific). Samples were homogenized in protein lysis buffer and all isolation was done on ice.

2.4.4 Statistical analysis:

Significance analysis of microarray (SAM) method was used to identify those proteins and phosphoproteins that were differentially expressed in ER-positive and ER-negative breast tumors. Analysis revealed a significant difference (permutation *P*-value< 0.01, hereafter referred to as *P*-value) in the expression of 40 proteins between ER-negative and ER-positive tumors with a false discovery rate (FDR) of 1%. Group characteristics were tabulated and compared between groups with the χ^2 test or

Kruskal-Wallis test as appropriate. All patients in the datasets with reliable clinical follow-up data were used for the outcome analyses. Overall survival (OS) was measured from the date of diagnosis to the date of death from any cause. Recurrence-free survival (RFS) was measured from the date of diagnosis to the date of breast cancer recurrence. Patients who died before experiencing a disease recurrence were considered censored at their date of death. Survival outcomes were estimated with the Kaplan-Meier method and compared between groups with the log-rank statistic. Multivariable Cox proportional hazards models were fit to determine the association of breast cancer subtypes with survival outcomes after adjustment for other patient characteristics. This analysis was done by Corey Speers with input from Dr. Susan Hilsenbeck at Baylor College of Medicine.

2.4.5 Kaplan-Meier Survival Analysis:

After hierarchical clustering to identify subgroups of ER-negative breast cancer, figure of merit scoring showed that these four clusters were stable against reclustering. Using this information each tumor sample was classified as falling into one of the 4 clusters or subgroups (ER-low, stathmin, S6 kinase, or HER2/neu subgroup of tumors). After classification of tumors, Kaplan-Meier analysis using the survival data from the training and validation set was done using GraphPad Prism 5.0 and survival curves were generated. Chi squared scores were calculated to determine significance and hazard

ratios (HR) and 95% confidence intervals (CI) we also calculated. This analysis was done by Corey Speers.

2.4.6 Censured Survival Analysis:

To identify proteins associated with good and poor outcomes we performed significance analysis for microarrays (SAM) using censured survival data [21]. Row average was used in the input engine and 250 permutations were run. S0 was calculated using the Tuscher *et al*. method as described previously [21]. Q-values were not calculated and the false discovery rate was <5%. Hierarchical clustering was done using all significant proteins and clustering was done using Pearson's correlation with complete linkage [22]. This analysis was done by Corey Speers.

2.4.7 Kruskal-Wallis Test:

The non-parametric Kruskal-Wallis test (an extension of the Mann-Whitney U test for more than 3 groups) was used to identify proteins associated with the groups defined by gene expression profiling limited to the intrinsic gene list. Using the intrinsic gene list previously defined, we were able to classify all of the training set tumors into a luminal A, luminal B, normal, erbB2, or basal like cluster [23-25]. Using these cluster definitions based on gene expression analysis allowed us to use Kruskal-Wallis test to identify proteins associated with these previously defined subtypes. Significance was based on estimated (Benjamini-Hochberg) and the selected false discover rate limit was

0.03. In this analysis, 47 samples were identified as being luminal A, 28 samples as luminal B, 42 samples as erbB2, 24 samples as normal, and 25 samples as basal. This analysis was done by Corey Speers with input from Dr. Susan Hilsenbeck at Baylor College of Medicine.

2.4.8 RNA and Protein Correlation:

RNA expression values were determined using the Applied Biosystem Human Genome Survey Microarray version 2.0 (Applied Biosystems, Foster City, CA). Data was normalized using QC-RMA and log2 transformed. RNA expression values were then compared to the corresponding protein expression values determined in the RPPA analysis. Comparisons were only done using total protein measurements as phosphoproteins and cleavage products were unsuitable for comparison with RNA expression products. Correlation was determined using Spearman's rank correlation. This is a distribution free, two sided test of independence between two variables. R values were calculated and *P*-values were determined using Gaussian approximation. This analysis was done by Corey Speers with input from Dr. Susan Hilsenbeck at Baylor College of Medicine.

2.5 The Materials and Methods of Chapter 5.

2.5.1 RNA isolation and Quantitative RT-PCR (Q-RT-PCR):

Total RNA was isolated using the RNeasy RNA isolation kit (QIAGEN, Valencia, CA) as described in the general methods section of this chapter. Briefly, quantitative RT-PCR (Q-RT-PCR) assays of transcripts were carried out using gene-specific double fluorescence-labeled probes in an ABI PRISM 7700 Sequence Detector (Applied Biosystem, Foster City, CA). The PCR reaction mixture and cycling conditions were as previously described. Primers were designed based on the Entrez ID NM_014791 for human maternal embryonic leucine zipper kinase (MELK) and were as follows:

Forward: ccaacaaaatattcatggttcttg

Reverse: aggcgatcctgggaaattat

Amplicon: caacaaaatattcatggttcttgagtactgccctggaggagagctgtttgactatataatttcccaggatcgcct with probe ID #25 from the human universal probe set library used (Roche Applied Science, Indianapolis, IN).

2.5.2 Western blot analysis:

Protein was isolated from breast cancer cell lines and human breast tumors as described in the general methods section of this chapter. Western blots were done as previously described using the primary total MELK antibody (Cell Signaling Technologies, Danvers, MA, catalog #2274) diluted 1:500. After washing in TBST, the membrane was

then incubated with horseradish peroxidase-conjugated secondary anti-rabbit antibody, washed again, and antigen-antibody complexes were detected using the ECL or ECL Plus chemiluminescent system (Amersham Bioscience). Quantitation was done by measuring pixel density of the MELK bands compared to loading control (beta actin) pixel intensity.

2.5.3 siRNA transfection:

siRNAs for MELK was purchased from Sigma Aldrich (St. Louis, MO). siRNA transfection was performed as described in the general methods section of this chapter. All cell lines used in the siRNA knockdown experiments were grown as previously described according to ATCC specifications (http://www.atcc.org).

2.5.4 Cell proliferation assays:

Cell growth was measured using the CellTiter 96^{TM} Aqueous Non-Radioactive Cell Proliferation assay (MTS assay, Promega) according to the manufacturer's instructions and as previously described in the general methods section of this chapter. Briefly, cells were plated in 96-well plates at 2000 cells per well. Every 24 hours, a solution containing 20:1 ratio of MTS and PMS was added to the cells and absorption was measured. Each data point was performed in heptuplicate, and the results were reported as average absorption ± standard deviation. The data is reported as percentage of growth compared to mock transfected controls for each cell line.

Experiments were repeated at least twice and the percentage growth inhibition is the average of the experiments.

2.5.5 Kaplan-Meier Survival Analysis:

Multiple available datasets were interrogated to evaluate the prognostic import of MELK in the Wang, Desmedt, van de Vijver, Ivshina, and Denmark datasets [1-4]. Data was obtained from the repository on gene expression omnibus (GEO) website at http://www.ncbi.nlm.nih.gov/geo/. Probe set IDs corresponding to MELK were extracted from the datasets, as were the corresponding annotation and clinical data. Datasets were sorted according to expression of MELK, from highest to lowest, Kaplan-Meier analysis using the survival data from the datasets was performed. Visualization and statistical testing was done using GraphPad Prism 5.0 and survival curves were generated. Chi squared scores were calculated to determine significance and hazard ratios (HR) and 95% confidence intervals (CI) were also calculated. Dataset categorization was done two ways; first by depicting higher and lower than mean MELK expression in the datasets, and second, by depicting quartile expression of MELK as listed in the figures in **Chapter 5**. This analysis was done by Corey Speers.

2.5.6 Multivariate Analysis:

A Cox proportional hazards model was constructed to identify potential factors of survival. Survival time was defined as the time from the date of the initial diagnosis

of breast cancer to the date of death. Subjects who were alive at the last follow-up were censored on that date. The initial multivariate model contained all variables independently associated with survival in univariate analyses. A backward selection procedure, using $p \leq 0.05$ as the criterion for inclusion, was then implemented to build the final model. All statistical tests were conducted using STATA 10.0 (College Station, TX) and were done by Krystal Sexton at Baylor College of Medicine.

2.6 Summary

This chapter has been a brief discussion of the most relevant materials and methods used to generate the data for this thesis. Additional information may be obtained by reviewing the laboratory notebooks from these studies archived under the supervision of Dr. Powel Brown. The following chapters will detail the studies conducted in this thesis as well as the results and summary of these findings.

References:

1. Wang, Y., et al., *Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.* Lancet, 2005. **365**(9460): p. 671-9.
2. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.
3. Desmedt, C., et al., *Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series.* Clin Cancer Res, 2007. **13**(11): p. 3207-14.
4. Ivshina, A.V., et al., *Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer.* Cancer Res, 2006. **66**(21): p. 10292-301.
5. Chang, J., et al., *Survival of patients with metastatic breast carcinoma: importance of prognostic markers of the primary tumor.* Cancer, 2003. **97**(3): p. 545-53.
6. Tham, Y.L., et al., *Clinical response to neoadjuvant docetaxel predicts improved outcome in patients with large locally advanced breast cancers.* Breast Cancer Res Treat, 2005. **94**(3): p. 279-84.
7. Cleator, S., et al., *Gene expression patterns for doxorubicin (Adriamycin) and cyclophosphamide (cytoxan) (AC) response and resistance.* Breast Cancer Res Treat, 2006. **95**(3): p. 229-33.
8. Mohsin, S.K., et al., *Neoadjuvant trastuzumab induces apoptosis in primary breast cancers.* J Clin Oncol, 2005. **23**(11): p. 2460-8.
9. Chang, J.C., et al., *Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer.* Lancet, 2003. **362**(9381): p. 362-9.
10. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.
11. Chan, E.Y., S. Kir, and S.A. Tooze, *siRNA screening of the kinome identifies ULK1 as a multidomain modulator of autophagy.* J Biol Chem, 2007. **282**(35): p. 25464-74.
12. Draviam, V.M., et al., *A functional genomic screen identifies a role for TAO1 kinase in spindle-checkpoint signalling.* Nat Cell Biol, 2007. **9**(5): p. 556-64.
13. Li, C. and W.H. Wong, *Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection.* Proc Natl Acad Sci U S A, 2001. **98**(1): p. 31-6.
14. Miura, K., et al., *Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles.* Cancer Res, 2002. **62**(11): p. 3244-50.
15. Zhang, B., et al., *GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies.* BMC Bioinformatics, 2004. **5**: p. 16.

16. Dalmasso, C., P. Broet, and T. Moreau, *A simple procedure for estimating the false discovery rate.* Bioinformatics, 2005. **21**(5): p. 660-8.
17. Fu, W.J., R.J. Carroll, and S. Wang, *Estimating misclassification error with small samples via bootstrap cross-validation.* Bioinformatics, 2005. **21**(9): p. 1979-86.
18. Whitlock, M.C., *Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach.* J Evol Biol, 2005. **18**(5): p. 1368-73.
19. Hu, J., et al., *Non-parametric quantification of protein lysate arrays.* Bioinformatics, 2007. **23**(15): p. 1986-94.
20. Lu, C., et al., *Effect of epidermal growth factor receptor inhibitor on development of estrogen receptor-negative mammary tumors.* J Natl Cancer Inst, 2003. **95**(24): p. 1825-33.
21. Tusher, V.G., R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response.* Proc Natl Acad Sci U S A, 2001. **98**(9): p. 5116-21.
22. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns.* Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14863-8.
23. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
24. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.
25. Sotiriou, C., et al., *Breast cancer classification and prognosis based on gene expression profiles from a population-based study.* Proc Natl Acad Sci U S A, 2003. **100**(18): p. 10393-8.

Chapter 3

Identification of Novel Kinase Targets for the Treatment of Estrogen Receptor-Negative Breast Cancer

3.1 Abstract

Previous gene expression profiling studies of breast cancer have focused on the entire genome to identify genes differentially expressed between estrogen receptor alpha (ER)-positive and ER-alpha-negative cancers. Here we used gene expression microarray profiling to identify a distinct kinase gene expression profile that identifies ER-negative breast tumors and subsets ER-negative breast tumors into 4 distinct subtypes. Based upon the types of kinases expressed in these clusters, we identify a cell cycle regulatory subset, a S6 kinase pathway cluster, an immunomodulatory kinase expressing cluster, and a MAPK pathway cluster. Furthermore, we show that this specific kinase profile is validated using independent sets of human tumors, and is also seen in a panel of breast cancer cell lines. Kinase expression knockdown studies show that many of these kinases are essential for the growth of ER-negative, but not ER-positive, breast cancer cell lines. Finally, survival analysis of patients with breast cancer shows that the S6 kinase pathway signature subtype of ER-negative cancers confers an extremely poor prognosis, while patients whose tumors express high levels of immunomodulatory kinases have a significantly better prognosis. This study identifies a list of kinases that are prognostic and may serve as druggable targets for the treatment of ER-negative breast cancer.

3.2 Introduction

The genomic era has produced an exponential increase in our understanding of cancer biology and has greatly accelerated cancer drug development. With the advent and implementation of microarray expression profiling, it is now possible to evaluate gene expression in tumors on a genome-wide basis. Gene expression analysis is now extensively used to subtype cancers, predict prognosis and disease free survival, and determine optimal treatment [1-7].

Estrogen receptor alpha (ER)-positive breast cancers account for 60-70% of breast cancers, but the remaining 30-40% of breast cancers are ER-negative and are poorly responsive to traditional therapies [8]. Selective estrogen receptor modulators (SERMs), such as tamoxifen and raloxifene, and aromatase inhibitors are currently used to treat ER-positive breast cancer and have been shown to reduce ER-positive breast cancer recurrence by approximately 50% [9]. These agents, however, are not effective in treating ER-negative breast cancer. Currently, chemotherapy is used to treat ER-negative tumors [10]. Such therapy is generally toxic and is not specifically targeted to ER-negative breast cancer.

A major goal of current breast cancer research has been to identify targets that are unique to cancer cells and to identify drugs that kill only cancerous cells without affecting normal tissue. While achieving this goal has been difficult, there are several examples of effective targeted therapies, including development of the monoclonal antibodies trastuzumab (targeting the HER2/neu receptor) and bevacizumab (targeting

vascular epithelial growth factor) which have been shown to be effective in treating breast cancer [11, 12]. Other successes include the development of small molecule tyrosine kinase inhibitors including gefitinib and erlotinib (both of which target the epidermal growth factor receptor), and lapatinib (a dual kinase inhibitor targeting both the epidermal growth factor receptor and the HER2/neu receptor) [13-16]. Despite these advances, such therapies are effective only in the 10-15% of patients whose tumors overexpress HER2. To develop targeted therapies for the remaining ER-negative breast cancers, including the aggressive ER-negative, PR-negative, HER2-negative ("triple-negative") breast cancers, we have used expression microarray analysis to identify molecules that play a role in breast cancer development and progression. Subsequent validation of these findings, along with the development of specific targeted inhibitors of these molecules, will certainly broaden treatment options and improve patient survival.

The purpose of this study was to identify the kinases that are over-expressed in ER-negative breast cancer and which may serve as "druggable targets" for the treatment of ER-negative breast cancer and in particular, "triple-negative" breast cancer. We have used transcriptional profiling data to evaluate the expression of the human kinome and have identified a set of kinases which are differentially expressed and are critical for the growth of ER-negative breast cancer. Our results also demonstrate that ER-negative breast cancer can be subdivided into four separate subgroups, each of which is distinct in the type and level of kinases they express. We

used siRNA knockdown studies to identify a subset of these kinases that are required for the growth of ER-negative breast cancer cells. These kinases represent promising targets for the treatment of ER-negative breast cancers.

3.3 Results

To identify kinases that are differentially expressed in ER-negative breast cancers, we designed a study to compare kinase expression levels in ER-positive and ER-negative human breast tumor samples. A summary of the study design is outlined in **Figure 3.1**.

3.3.1 Patient Population

A total of 102 patients with invasive breast cancer were recruited by Dr. Jenny Chang through IRB-approved, neoadjuvant studies to investigate gene expression in human tumors before and after drug treatment. Breast biopsies using a core needle were taken by Dr. Jenny Chang before initiation of any treatment and were used in this study. Because the patients did not receive systemic adjuvant or neoadjuvant therapy prior to the biopsy, the results from the gene expression analysis represent basal gene expression in these breast cancers. For these gene expression profiling experiments, 102 breast tumors were studied in the laboratory of Dr. Jenny Chang, 58 of which were ER-positive and 44 ER-negative by IHC-staining (24 of which were confirmed as "triple-negative"). The tumors were all stage III or IV from pre- and post-menopausal women, with all tumors showing >30% cellularity. The women were from several racial groups (as shown in **Table 3.1**) and the majority had no palpable nodes at baseline. Most of the women were premenopausal and presented with relatively large tumors (ranging from

Figure 3.1 Overview of Study Design

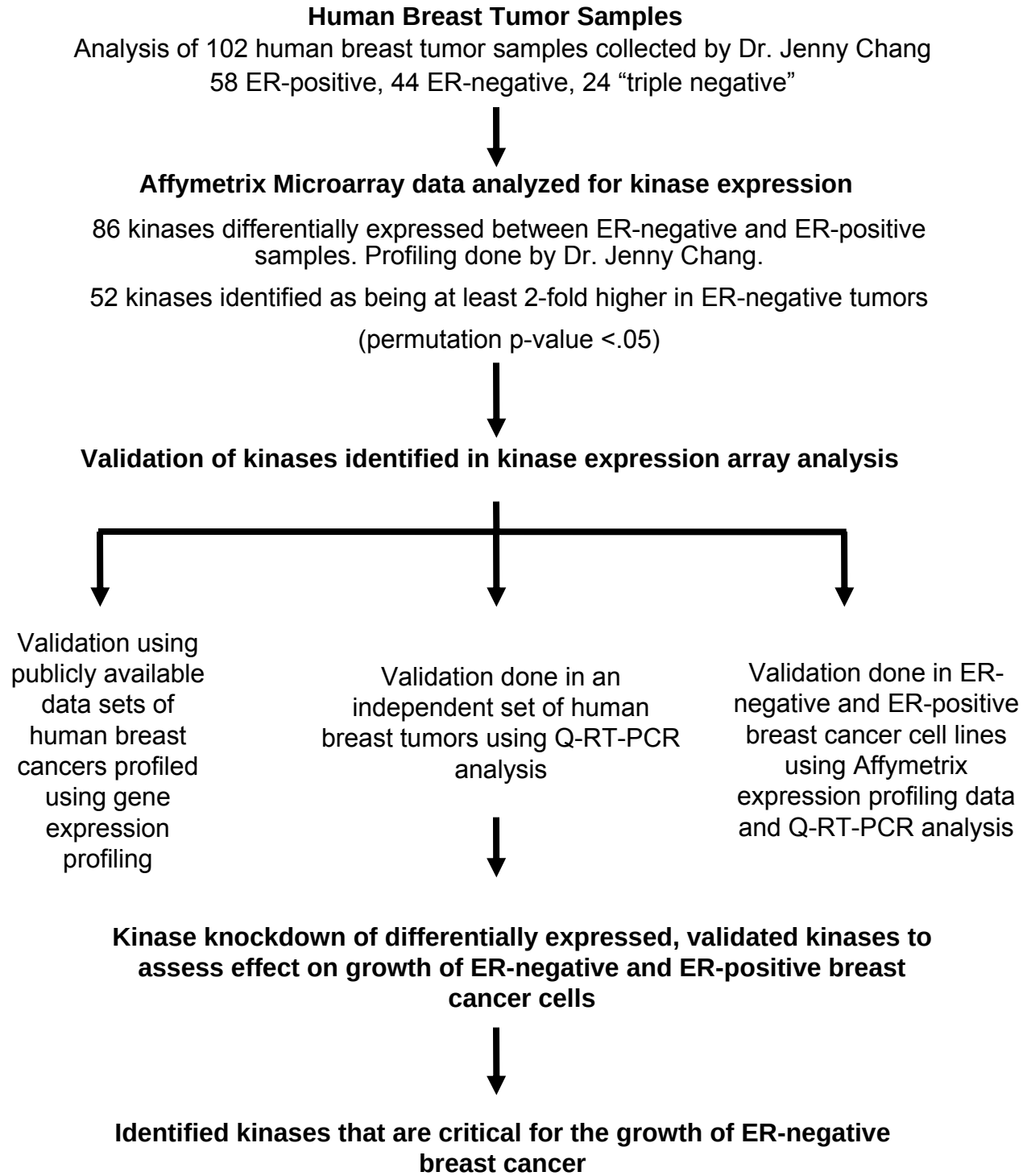


Table 3.1. Clinical characteristics of the patients and tumor samples used in the study.

Characteristic	Tumor Set N=102 (%)
Age	
Mean	48.1
Range	(32-72)
Race	
Caucasian	50 (57%)
Hispanic	7 (8%)
African-American	23 (27%)
Asian	7 (8%)
Menopausal Status	
Pre	49 (62%)
Post	30 (38%)
BMI	
Mean	29.7
Range	(16.1-48.3)
Baseline Tumor Size, cm	
Mean	6.3
Range	(2.5-25.0)
Palpable Nodes at Baseline	
Yes	20 (21%)
No	77 (79%)
ER	
Positive	57 (56%)
Negative	45 (44%)
Unknown	0 (0%)
PR	
Positive	37 (36%)
Negative	47 (46%)
Unknown	18 (18%)
HER2/Neu	
Positive	27 (26%)
Negative	58 (57%)
Unknown	17 (17%)

Table 3.1- Characteristics of 102 patients with breast cancer. Tumors from these patients were collected by Dr. Jenny Chang and used for gene expression profiling to identify overexpressed kinases in ER-negative breast tumors

2.5 to 25 cm). The clinical and demographic features of these tumors are summarized in **Table 3.1**.

3.3.2 Affymetrix Gene Expression Profiling Identified Kinases Overexpressed in Human ER-Negative Breast Tumors

To identify signaling molecules that are differentially expressed in ER-negative breast cancers, we performed Affymetrix gene expression profiling to compare human ER-negative and ER-positive breast tumors. The profiling was done in the laboratory of Dr. Jenny Chang, with data generously provided for further analysis. Subsequent data analysis and clustering was limited to the known kinome with interrogation of the 779 known and putative human protein, nucleotide, and lipid kinases as well as kinase-interacting proteins and regulatory subunits as previously described [17-19]. These kinases and kinase associated genes are listed in **Supplementary Table 3.1**. We first performed analysis to identify those kinases that were differentially expressed in ER-positive and ER-negative breast tumors. Our analysis revealed a significant difference (permutation *P*-value< 0.05, hereafter referred to as *P*-value) in the expression of 86 kinases between ER-negative and ER-positive tumors, with a false discovery rate (FDR) of less than 1% (**Tables 3.2 and 3.3**). To visualize the clustering of the ER-positive and ER-negative tumors, hierarchical clustering analysis was done using only those kinases identified as being differentially expressed between the two groups (**Figure 3.2A**). Hierarchical clustering showed that these 86 kinases were able to segregate ER-positive

Table 3.2:
Kinases overexpressed in ER-negative breast cancer

Gene name	EntrezGene ID	fold change over ER-positive	perm. p-value
PFKP: phosphofructokinase, platelet	5214	3.84	0.000007
CXCL10: chemokine (C-X-C motif) ligand 10	3627	3.81	0.000026
MET	4233	2.75	0.000052
maternal embryonic leucine zipper kinase	9833	2.72	0.000005
PDXK	8566	2.70	0.000001
LYN	4067	2.60	0.000114
CCL4: chemokine (C-C motif) ligand 4	6351	2.54	0.000298
CHEK1	1111	2.54	0.000049
SRPK1	6732	2.52	0.000550
EGFR	1956	2.51	0.000112
PRKX /// PRKY	5613	2.49	0.000009
RIPK4	54101	2.48	0.000015
AURKB: aurora kinase B	9212	2.48	0.000571
BUB1	699	2.39	0.000013
YES1	7525	2.37	0.000022
LCK	3932	2.31	0.000358
SEPHS1: selenophosphate synthetase 1	22929	2.28	0.000173
CDC2	983	2.27	0.000311
UGP2	7360	2.24	0.000002
SGK	6446	2.22	0.000042
LYN	4067	2.22	0.000160
CHEK1	1111	2.22	0.000022
MAP4K4	9448	2.22	0.000077
PLK1	5347	2.21	0.005490
CCL2: chemokine (C-C motif) ligand 2	6347	2.19	0.006306
IRAK1	3654	2.16	0.000138
PTK7	5754	2.16	0.002156
RPS6KA1	6195	2.15	0.000423
PIM1: pim-1 oncogene	5292	2.15	0.000283
MPZL1: myelin protein zero-like 1	644387	2.13	0.000000
EPHA2	1969	2.12	0.000126
CDC7: cell division cycle 7 homolog	8317	2.10	0.002258
STK38L: serine/threonine kinase 38 like	23012	2.10	0.000754
SMG1	23049	2.10	0.000520
RIOK3	8780	2.09	0.000243
PGK1: phosphoglycerate kinase 1	5230	2.09	0.004493
PRKX: protein kinase, X-linked	5613	2.09	0.000768
YWHAQ	10971	2.07	0.000049
YES1	7525	2.07	0.000002
LYN	4067	2.07	0.000030
STK38: serine/threonine kinase 38	11329	2.06	0.000093
MAP3K5	4217	2.06	0.002712
PIK3CB	5291	2.05	0.000158
EPHB2	2048	2.05	0.000216

Table 3.2 (continued):
Kinases overexpressed in ER-negative breast cancer

Gene name	EntrezGene ID	fold change over ER-positive	perm. p-value
MAP4K2	5871	2.05	0.005400
SRPK1: SFRS protein kinase 1	6732	2.04	0.000081
VRK2	7444	2.04	0.000158
CSK	1445	2.04	0.000545
DAPK1: death-associated protein kinase 1	1612	2.03	0.000109
MALT1	10892	2.03	0.000043
SEPHS1: selenophosphate synthetase 1	22929	2.03	0.001312
UCK2: uridine-cytidine kinase 2	7371	2.03	0.000253
EPHB6	2051	2.02	0.002540
MAPK1	5594	2.01	0.001285
LIMK2	3985	2.00	0.000162
RYK: RYK receptor-like tyrosine kinase	6259	2.00	0.000191
RPS6KA3	6197	2.00	0.000657
PGK1: phosphoglycerate kinase 1	5230	2.00	0.000284
EPHB4: EPH receptor B4	2050	2.00	0.001583
TTK: TTK protein kinase	7272	2.00	0.000435
AK2: adenylate kinase 2	204	2.00	0.000063

Table 3.3:
Kinases overexpressed in ER-positive breast cancer

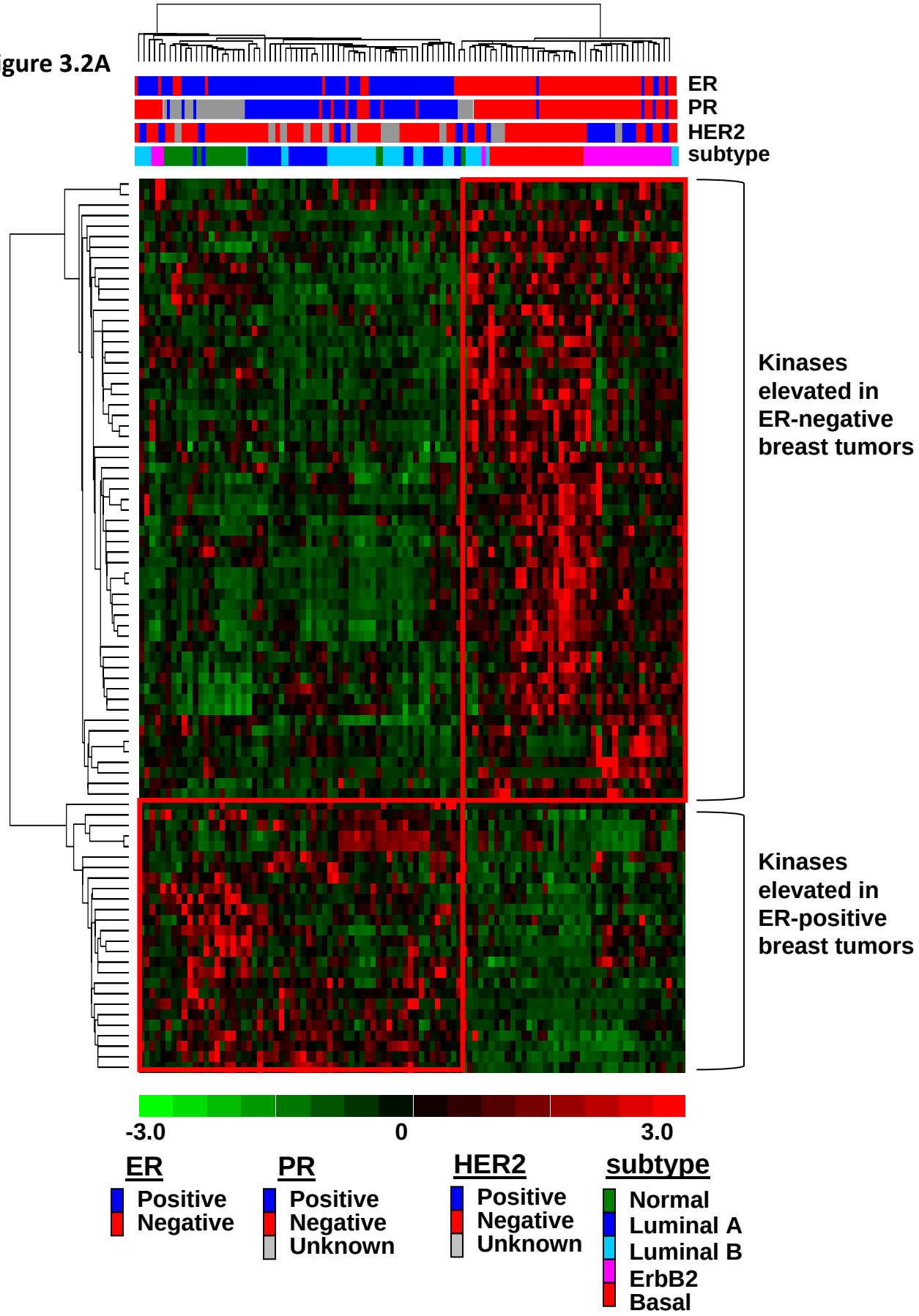
Affy probe set ID	Gene name	EntrezGene ID	fold change over ER-negative	perm. p-value
204379_s_at	FGFR3: fibroblast growth factor receptor 3	2261	4.32	0.000061
204686_at	IRS1: insulin receptor substrate 1	3667	3.47	0.00001
221667_s_at	HSPB8: heat shock 22kDa protein 8	26353	3.46	0.000085
204014_at	DUSP4: dual specificity phosphatase 4	1846	3.42	0.004515
220038_at	SGK3: serum/glucocorticoid regulated kinase family, member 3	23678	3.30	0.016929
211535_s_at	FGFR1: fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)	2260	2.80	0.019118
205399_at	DCLK1: doublecortin-like kinase 1	9201	2.74	0.014443
209341_s_at	IKBKB: inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	3551	2.70	0.000001
206197_at	NME5: non-metastatic cells 5, protein expressed in (nucleoside-diphosphate kinase)	8382	2.67	0.000016
202962_at	KIF13B: kinesin family member 13B	23303	2.58	0.000943
221207_s_at	NBEA: neurobeachin	26960	2.56	0.003208
202786_at	STK39: serine threonine kinase 39 (STE20/SPS1 homolog, yeast)	27347	2.53	0.000082
210740_s_at	ITPK1: inositol 1,3,4-triphosphate 5/6 kinase	3705	2.50	0.000034
207119_at	PRKG1: protein kinase, cGMP-dependent, type I	5592	2.42	0.015191
219686_at	STK32B: serine/threonine kinase 32B	55351	2.42	0.000189
205448_s_at	MAP3K12: mitogen-activated protein kinase kinase kinase 12	7786	2.32	0.000002
202454_s_at	ERBB3: v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	2065	2.29	0.000206
204862_s_at	NME3: non-metastatic cells 3, protein expressed in	4832	2.27	0.000627
207169_x_at	DDR1: discoidin domain receptor family, member 1	780	2.26	0.008385
210749_x_at	DDR1: discoidin domain receptor family, member 1	780	2.24	0.009653
208779_x_at	DDR1: discoidin domain receptor family, member 1	780	2.22	0.011492
221918_at	PCTK2: PCTAIRE protein kinase 2	5128	2.22	0.000091
208383_s_at	PCK1: phosphoenolpyruvate carboxykinase 1 (soluble)	5105	2.18	0.004562
213264_at	PCBP2: poly(rC) binding protein 2	5094	2.18	0.000343

Table 3.3 (continued):
Kinases overexpressed in ER-positive breast cancer

Affy probe set ID	Gene name	EntrezGene ID	fold change over ER-negative	perm. p-value
202281_at	GAK: cyclin G associated kinase	2580	2.14	0.009556
40225_at	GAK: cyclin G associated kinase	2580	2.13	0.012694
57540_at	RBKS: ribokinase	64080	2.11	0.000008
204589_at	NUAK1: NUAK family, SNF1-like kinase, 1	9891	2.11	0.018574

Figure 3.2A - Supervised hierarchical clustering identifies different subsets of ER-negative breast cancer. (**A**) Hierarchical clustering analysis of kinases that distinguish ER-positive from ER-negative human breast tumors. Gene expression analysis of 102 human breast tumors reveals 86 kinases that are differentially expressed between ER-negative and ER-positive human breast tumors with a permutation *P*-value <.05. Clustering done by Corey Speers with input from Dr. Anna Tsimelzon and Dr. Susan Hilsenbeck.

Figure 3.2A


ER
PR
HER2
subtype
Kinases elevated in ER-negative breast tumors
Kinases elevated in ER-positive breast tumors
-3.0
0
3.0
ER
Positive
Negative
PR
Positive
Negative
Unknown
HER2
Positive
Negative
Unknown
subtype
Normal
Luminal A
Luminal B
ErbB2
Basal

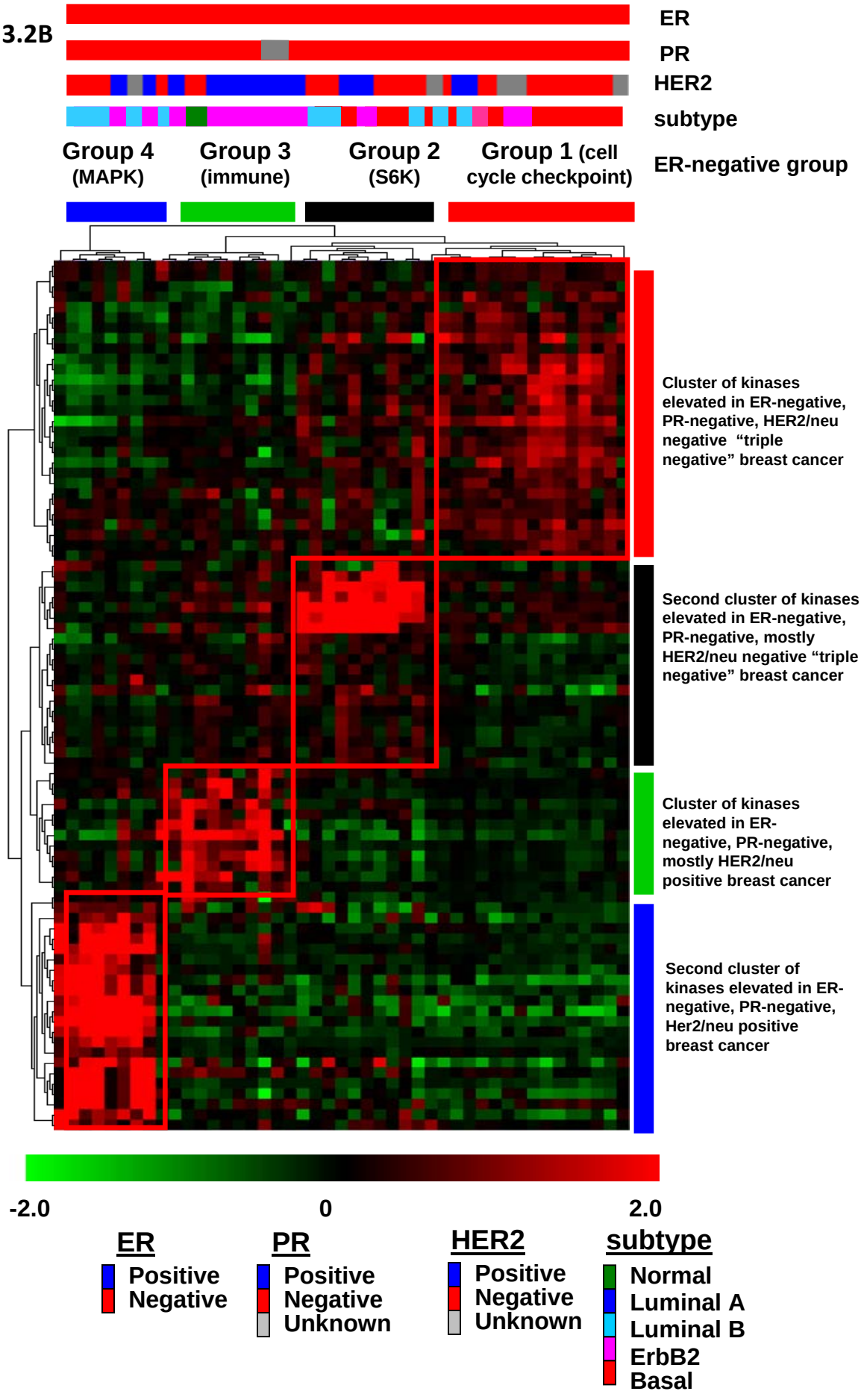
and ER-negative tumors and that the majority of the HER2-positive tumors were ER-negative (as expected). Upon further analysis, 52 of these 86 differentially expressed kinases were expressed at least 2 fold higher with a *P*-value <.05 in the ER-negative breast tumors as compared to ER-positive tumors. These 52 kinases were selected for further study. Original data analysis in dChip and with BRB array tools was done by Dr. Anna Tzsimelson and Dr. Susan Hilsenbeck. Reanalysis with MeV and dChip and further bioinformatic analysis was done by Corey Speers.

3.3.3 Unsupervised Clustering Analysis Revealed Four Distinct Subtypes of ER-negative Breast Cancer

We next determined whether this list of 52 kinases overexpressed in ER-negative breast cancers could subcluster the ER-negative tumors in an unbiased manner. We performed unsupervised clustering analysis using only the ER-negative breast cancer samples and found that these tumors clustered broadly into 4 distinct subtypes of ER-negative breast cancer **(Figure 3.2B**, labeled as groups 1-4). Upon further inspection of these four subsets of tumors, there was one subset of tumors defined by kinases involved in cell cycle checkpoint control and mitogenesis, including *CHK1*, *BUB1*, *TTK*, and *AK2* (group 1 termed the "cell cycle checkpoint group"). Another tumor subset was defined by kinases involved in the S6 kinase signaling pathway and includes *RPS6KA3, SMG-1*, and *RPS6KA1* kinases (group 2 termed the "S6 kinase group"). Of the two other ER-negative clusters, one is defined by kinases that are involved in modulating the

Figure 3.2B - Supervised hierarchical clustering identifies different subsets of ER-negative breast cancer. (**B**) Unsupervised hierarchical clustering analysis of overexpressed kinases only in ER-negative tumors using kinases overexpressed in ER-negative breast cancers reveals 4 distinct subsets of ER-negative breast cancer. These four subset are defined by kinases that are involved in cell cycle checkpoint control (group 1), S6 kinase signaling (group 2), immunomodulatory (group 3), or paracrine signaling involving many MAPKs (group 4). Subtype refers to the breast cancer subtypes identified by Sorlie *et al*. (6) Clustering done by Corey Speers with input from Dr. Anna Tsimelzon and Dr. Susan Hilsenbeck.

Figure 3.2B


ER
PR
HER2
subtype
Group 4 (MAPK)
Group 3 (immune)
Group 2 (S6K)
Group 1 (cell cycle checkpoint)
ER-negative group
Cluster of kinases elevated in ER-negative, PR-negative, HER2/neu negative "triple negative" breast cancer
Second cluster of kinases elevated in ER-negative, PR-negative, mostly HER2/neu negative "triple negative" breast cancer
Cluster of kinases elevated in ER-negative, PR-negative, mostly HER2/neu positive breast cancer
Second cluster of kinases elevated in ER-negative, PR-negative, Her2/neu positive breast cancer
-2.0
0
2.0
ER
Positive
Negative
PR
Positive
Negative
Unknown
HER2
Positive
Negative
Unknown
subtype
Normal
Luminal A
Luminal B
ErbB2
Basal

immune system (*IRAK1, TLR1, LCK*, and *LYN*) (group 3, termed the "immunomodulatory group"). The fourth group is defined by kinases that govern paracrine growth signaling and include mitogen activated protein kinases *MAP4K2, MAP4K4,* and *MAPK1* (group 4 termed the "MAPK group").

3.3.4 Gene Ontology Analysis

To gain insight into the function of kinases highly expressed in ER-negative breast cancer, we performed gene ontology (GO) enrichment analysis using EASE and found that several classes of biological function were highly enriched in our selected sets (**Table 3.4**). We observed enrichment for kinases involved in the regulation of metabolism (*P*-value $<10^{-14}$), cell cycle checkpoint control (*P*-value $<10^{-12}$), DNA damage checkpoint control (*P*-value $<10^{-11}$), cell-to-cell signaling (*P*-value $<10^{-9}$), and apoptosis regulation (*P*-value $<10^{-9}$). Many of these kinases fell in linear pathways, for example *TTK*, *CHK1*, and *BUB1* kinases (group 1), all of which play a role via sequential phosphorylation and activation in regulating G2/M transitioning as well as DNA damage checkpoint control pathways, and *PIK3CB, RPS6KA1*, and *SMG-1* kinases (group 2), that mark tumors with activated cytoplasmic kinases involved in mitogenesis and signal transduction.

Table 3.4. Kinases identified in analysis as most highly overexpressed in ER-negative tumors

Cell cycle checkpoint cluster	Gene bank accession number	Kinase function
BUB1	NM_001211	cell cycle checkpoint
CHK1 checkpoint homolog	NM_001274	cell cycle checkpoint
TTK protein kinase	NM_003318	cell cycle checkpoint
serum/glucocorticoid regulated kinase	NM_005627	cell cycle checkpoint
SFRS protein kinase 1	NM_003137	cell cycle checkpoint
maternal embryonic leucine zipper kinase	NM_014791	cell cycle checkpoint
RYK receptor-like tyrosine kinase	NM_001005861	positive regulation of proliferation
vaccinia related kinase	NM_006296	anti-apoptosis
phosphoglycerate kinase 1	NM_000291	metabolism
selenophosphate synthetase 1	NM_004226	metabolism
uridine-cytidine kinase 2	NM_012474	metabolism
UDP-glucose pyrophosphorylase 2	NM_006759	metabolism
adenylate kinase 2	NM_001625	metabolism
aurora kinase B	NM_004217	cell cycle checkpoint
cell division cycle 2	NM_001786	cell cycle checkpoint
cell division cycle 7 homolog	NM_003503	cell cycle checkpoint
S6 kinase pathway cluster		
ribosomal protein S6 kinase, 90kDa, polypeptide 1	NM_001006665	positive regulation of proliferation
PI-3-kinase-related kinase SMG-1	NM_015092	DNA repair
EPH receptor B4	NM_004444	positive regulation of proliferation
serine/threonine kinase 38 like (NDR2)	NM_015000	positive regulation of proliferation
PI3K catalytic subunit beta	NM_006219	positive regulation of proliferation
death-associated protein kinase 1	NM_004938	anti-apoptosis
pim-1 oncogene	NM_002648	anti-apoptosis
LIM domain kinase 2	NM_001031801	cell adhesion
phosphoribosyl pyrophosphate synthetase 1	NM_002764	metabolism
EPH receptor B6	D83492	nervous system development
EPH receptor B2	NM_017449	maintainance of polarity
ribosomal protein S6 kinase, 90kDa, polypeptide 3	NM_001006665	positive regulation of proliferation
MAPK cluster		
mitogen-activated protein 4K4	NM_004834	response to stress
mitogen-activated protein kinase kinase 6	NM_002758	DNA damage, cell cycle arrest
mitogen-activated protein kinase 1 (ERK2)	NM_002745	positive regulation of proliferation
mitogen-activated protein 4K2	NM_004579	positive regulation of proliferation
mindbomb homolog 1 (14-3-3)	NM_020774	receptor mediated endocytosis
v-raf-1 murine leukemia viral oncogene homolog	NM_002880	anti-apoptosis
protein kinase, X-linked	NM_005044	unknown
PTK7 protein tyrosine kinase 7	NM_002821	cell adhesion
myelin protein zero-like 1	NM_003953	cell to cell signaling
phosphofructokinase, platelet	NM_002627	metabolism
epidermal growth factor receptor	NM_005228	positive regulation of proliferation
MET proto-oncogene	NM_000245	activation of MAPK activity
Immunomodulatory cluster		
toll-like receptor 1	NM_003263	immune system modulation
MALT lymphoma translocation gene 1	NM_006785	anti-apoptosis
serine/threonine kinase 17b	NM_004226	anti-apoptosis
interleukin-1 receptor-associated kinase 1	NM_001569	positive regulation of transcription
chemokine (C-X-C motif) ligand 10	NM_001565	immune system modulation
lymphocyte-specific protein tyrosine kinase	NM_001042771	immune system modulation
chemokine (C-C motif) ligand 4	NM_002984	cell to cell signaling
pyridoxal (pyridoxine, vitamin B6) kinase	NM_003681	metabolism
v-yes-1	NM_005433	positive regulation of proliferation

Table 3.4- 52 overexpressed kinases in ER-negative breast cancer fall into 4 distinct subsets with varying biological functions. Gene ontology analysis shows that these kinases have varying biological functions, but most regulate growth, affect cell cycle, or are involved in DNA damage sensing and repair.

3.3.5 Differentially Expressed Kinases Validated Using Publicly Available Data Sets

Having identified differentially expressed kinases using gene expression profiling in a test set of tumors, we next wanted to validate these kinases as being differentially expressed in an independent set of human breast tumors. To demonstrate that these kinases are indeed overexpressed in ER-negative tumors compared to ER-positive tumors, we analyzed gene expression data from 12 additional publically available data sets to validate the list of kinases we found overexpressed in our tumors. This data set from multiple investigators includes over 1800 additional tumor samples (556 ER-negative and 1282 ER-positive tumors) for which there is gene expression profiling data and is the most comprehensive breast tumor set available [3, 20-29]. To utilize the power of such a large combined dataset, we employed a technique recently described by Whitlock that relies on a weighted Z-method to combine *P*-values [30]. This robust approach, superior to Fisher's combined probability test, revealed that all of the selected 52 kinases validated as being significantly more highly expressed (with extremely high z-scores and low *P*-values) in ER-negative breast tumors as compared to ER-positive tumors in an effective sample size of over 1800 tumors (**Table 3.5**).

3.3.6 Validation of Kinase Overexpression in an Independent Set of Human Breast Tumors

We also wanted to confirm that the overexpression identified using Affymetrix gene expression profiling could be validated using another technique. We therefore

Table 3.5- **Validation of kinases in 12 different breast tumor datasets**

Gene symbol	Z-score	*P* -value	Gene symbol	Z-score	*P* -value
SRPK1	20.84	< 1.0 e-20	**AK2**	14.97	< 1.0 e-20
PRKX	20.39	< 1.0 e-20	**UGP2**	14.76	< 1.0 e-20
MELK	19.76	< 1.0 e-20	**LIMK2**	14.52	< 1.0 e-20
SRPK1	19.72	< 1.0 e-20	**PIM1**	14.44	< 1.0 e-20
PFKP	19.69	< 1.0 e-20	**CDC2**	14.36	< 1.0 e-20
EGFR	19.59	< 1.0 e-20	**LCK**	14.14	< 1.0 e-20
LYN	19.57	< 1.0 e-20	**MET**	13.80	< 1.0 e-20
CHEK1	19.02	< 1.0 e-20	**EPHA2**	13.73	< 1.0 e-20
BUB1	18.64	< 1.0 e-20	**CCL2**	13.59	< 1.0 e-20
DAPK1	18.50	< 1.0 e-20	**AURKB**	13.49	< 1.0 e-20
TTK	18.14	< 1.0 e-20	**RPS6KA3**	13.25	< 1.0 e-20
YES1	17.86	< 1.0 e-20	**EPHB6**	13.03	< 1.0 e-20
MAP4K4	17.81	< 1.0 e-20	**MAPK1**	12.35	< 1.0 e-20
STK38	17.81	< 1.0 e-20	**MAP3K5**	12.30	< 1.0 e-20
RIPK4	17.43	< 1.0 e-20	**EPHB2**	12.20	< 1.0 e-20
UCK2	17.02	< 1.0 e-20	**EPHB4**	12.19	< 1.0 e-20
CXCL10	16.87	< 1.0 e-20	**CSK**	12.18	< 1.0 e-20
PDXK	16.87	< 1.0 e-20	**CCL4**	10.86	< 1.0 e-20
SEPHS1	16.55	< 1.0 e-20	**STK38L**	10.80	< 1.0 e-20
IRAK1	16.51	< 1.0 e-20	**SGK**	10.40	< 1.0 e-20
MALT1	16.43	< 1.0 e-20	**RIOK3**	9.93	2.2 e-20
PLK1	15.89	< 1.0 e-20	**VRK2**	8.90	1.0 e-17
YWHAQ	15.46	< 1.0 e-20	**RPS6KA1**	5.51	0.000000005
PGK1	15.20	< 1.0 e-20	**MAP4K2**	5.15	0.00000008
CDC7	15.19	< 1.0 e-20	**RYK**	4.55	0.0000008
PTK7	15.16	< 1.0 e-20	**PIK3CB**	4.31	0.000009

Table 3.5- Kinase overexpression validated in independent human tumor sample data sets.
Data analysis of an additional 12 publically available human breast tumor datasets shows the 52 kinases identified as being overexpressed in this study are also significantly overexpressed in ER-negative breast tumors in the other datasets. Z-scores were calculated using the Z-transform test and are listed with their correlating *P* -value. Analysis performed by Corey Speers with input from Dr. Susan Hilsenbeck.

used an independent set of 60 human breast tumors from the tumor bank at Baylor College of Medicine for further validation. After identifying approximately equal numbers of ER-positive and negative samples, we used quantitative RT-PCR (Q-RT-PCR) to confirm the overexpression of the kinases identified in the array profiling. To date, 34 of the 34 kinases assayed were significantly more highly expressed in ER-negative human breast tumors than ER-positive tumors in this additional tumor set (*P*-value <0.05). Representative results from these experiments showing expression of six kinases (*CHK1, BUB1, PTK7, TTK1, TLR1*, and *RAF1*) are displayed in **Figure 3.3**.

3.3.7 Validation of Kinase Overexpression in Breast Cancer Cell Lines

To conduct further *in vitro* experimentation in cell lines, we wanted to confirm that our selected kinases identified in human breast tumors were also overexpressed in ER-negative breast cancer cell lines. Twelve ER-positive or ER-negative breast cancer cell lines were chosen and the expression of the identified kinases was measured under basal growth conditions. Of 42 kinases evaluated to date, all 42 were found to be statistically significantly increased (*P*-value <.05) in this panel of ER-negative breast cancer cell lines as compared to ER-positive cell lines using Q-RT-PCR. Representative results for several of these kinases (*CHK1, BUB1, PTK7, TTK1, TLR1*, and *RAF1*) are shown in **Figure 3.4**.

Figure 3.3 - Kinase overexpression validated in a panel of human tumor samples. The expression of 34 of 34 kinases identified in the array profiling were validated as being more highly expressed in ER-negative tumors compared to ER-positive tumors as measured by Q-RT-PCR in an independent set of breast tumors. Expression data for 6 representative kinases (*CHK1, BUB1, PTK7, TTK, TLR1*, and *RAF1*) are shown. Asterisks indicate *P*-value <0.01. Data are represented as mean ± SEM.

Figure 3.3

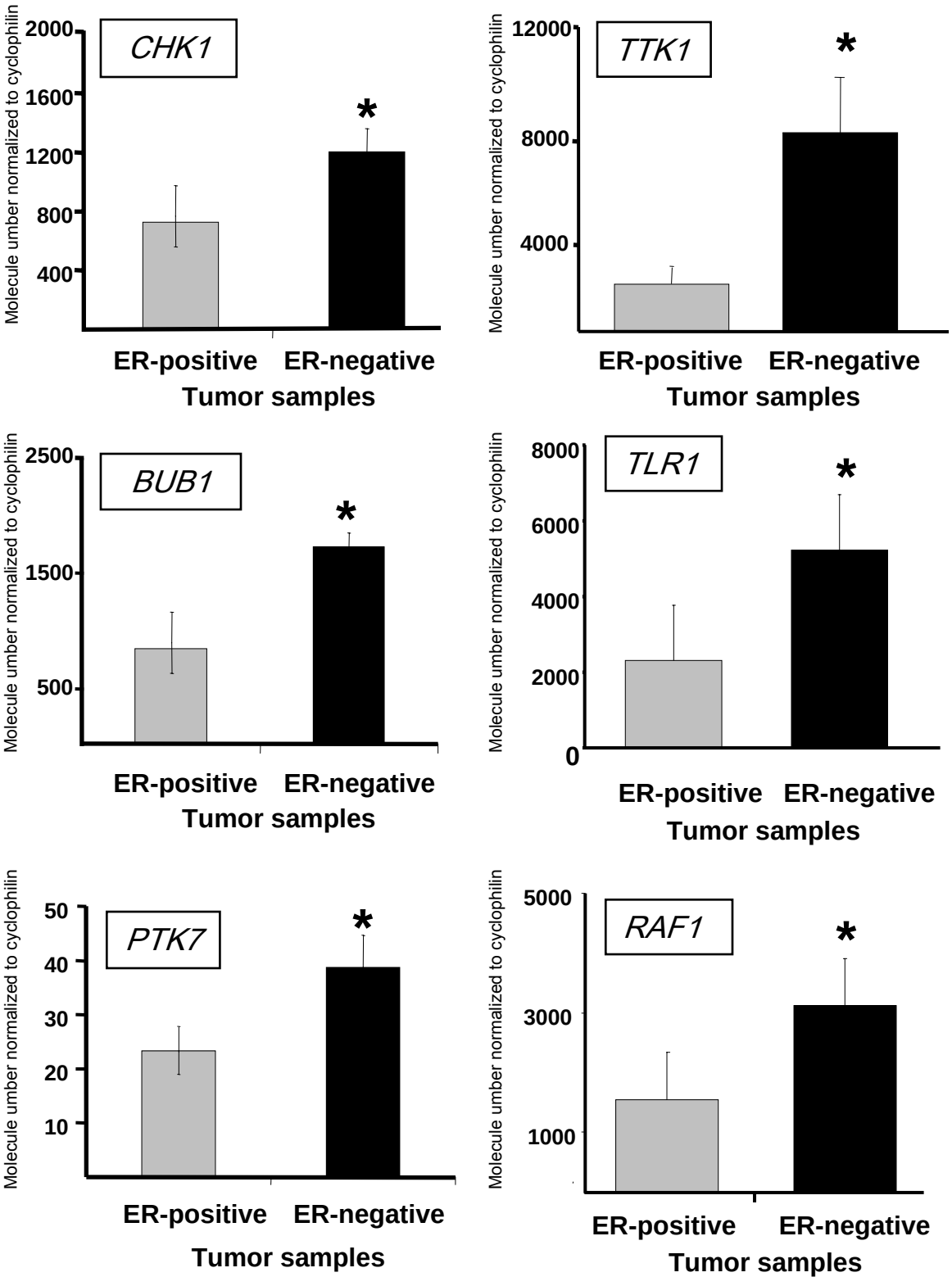
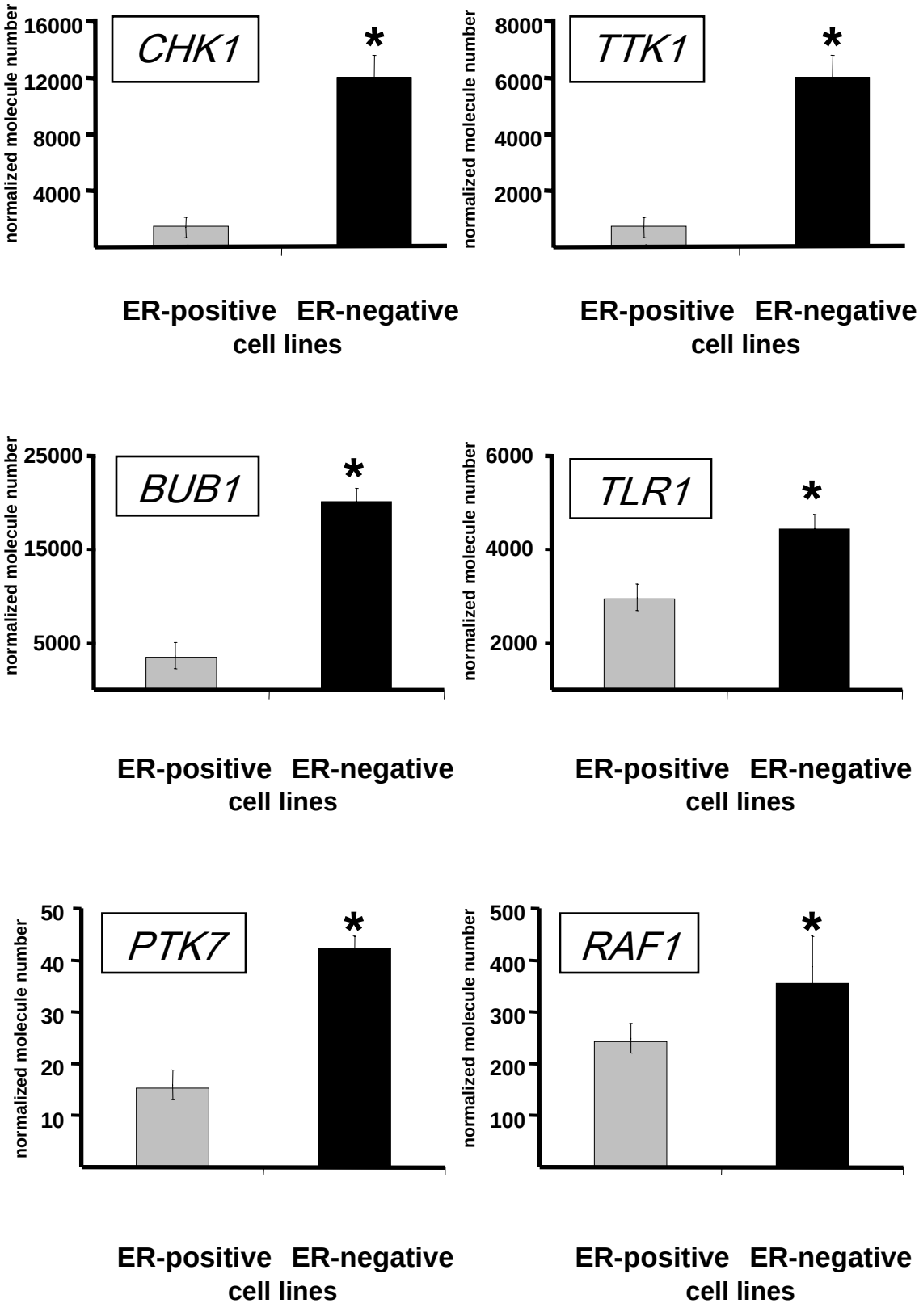

CHK1
Molecule umber normalized to cyclophilin
2000
1600
1200
800
400
*
ER-positive
ER-negative
Tumor samples
TTK1
Molecule umber normalized to cyclophilin
12000
8000
4000
*
ER-positive
ER-negative
Tumor samples
BUB1
Molecule umber normalized to cyclophilin
2500
1500
500
*
ER-positive
ER-negative
Tumor samples
TLR1
Molecule umber normalized to cyclophilin
8000
6000
4000
2000
0
*
ER-positive
ER-negative
Tumor samples
PTK7
Molecule umber normalized to cyclophilin
50
40
30
20
10
*
ER-positive
ER-negative
Tumor samples
RAF1
Molecule umber normalized to cyclophilin
5000
3000
1000
*
ER-positive
ER-negative
Tumor samples

Figure 3.4 - Kinase overexpression validated in a panel of breast cancer cell lines. The expression of 42 of 42 kinases was significantly higher in ER-negative breast cancer cell lines as compared to ER-positive cell lines. Again, expression data as measured by Q-RT-PCR, this time in a panel of breast cancer cell lines, for 6 representative kinases (*CHK1, BUB1, PTK7, TTK, TLR1*, and *RAF1*) are shown. Asterisks indicate *P*-value <0.01. Data are represented as mean ± SEM.

Figure 3.4

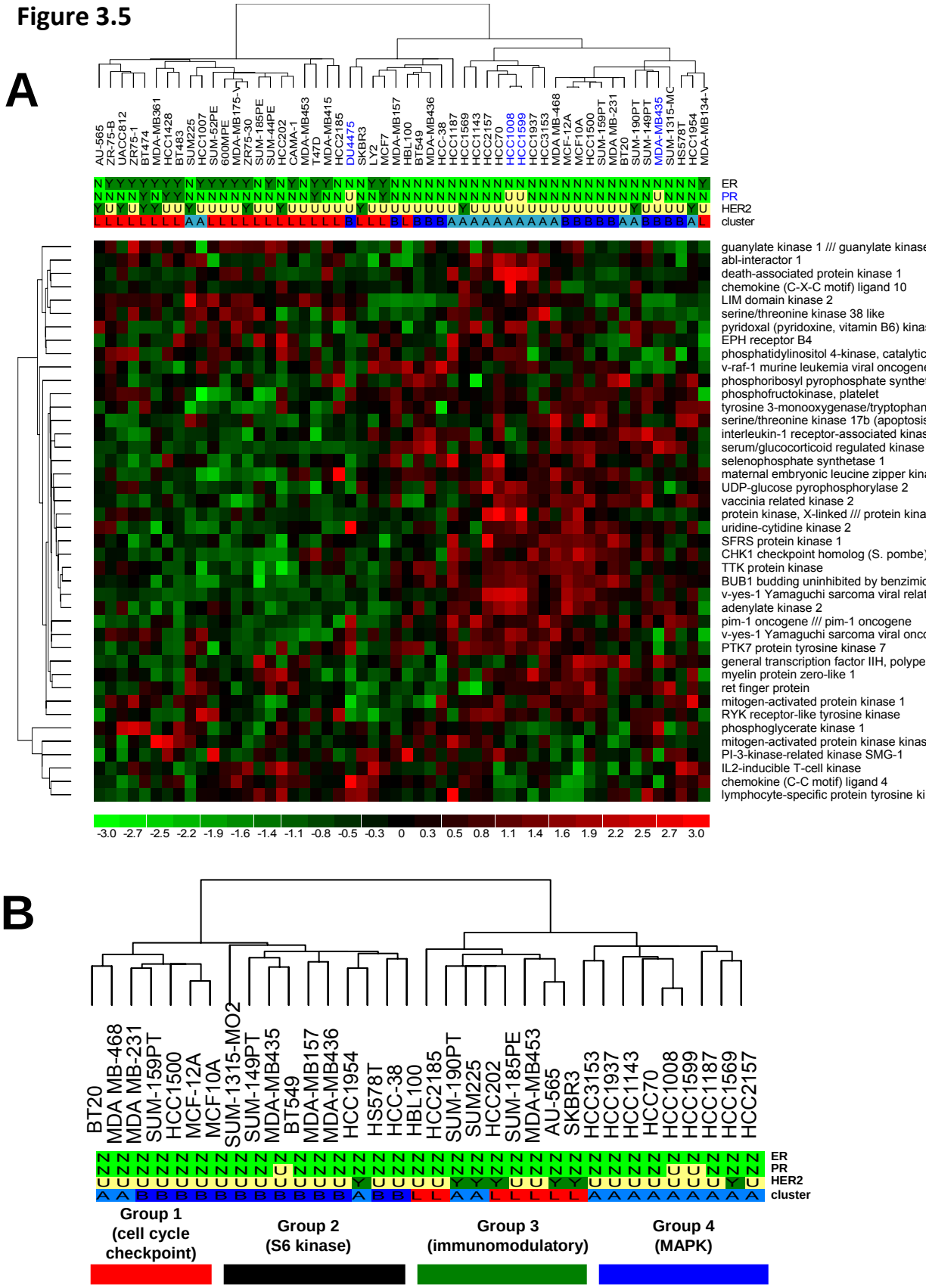

CHK1
normalized molecule number
16000
12000
8000
4000
*
ER-positive ER-negative
cell lines
TTK1
normalized molecule number
8000
6000
4000
2000
*
ER-positive ER-negative
cell lines
BUB1
normalized molecule number
25000
15000
5000
*
ER-positive ER-negative
cell lines
TLR1
normalized molecule number
6000
4000
2000
*
ER-positive ER-negative
cell lines
PTK7
normalized molecule number
50
40
30
20
10
*
ER-positive ER-negative
cell lines
RAF1
normalized molecule number
500
400
300
200
100
*
ER-positive ER-negative
cell lines

3.3.8 Cluster of Human Breast Cancer Cell Lines using the Kinase Profile

To determine whether the 52 kinases could accurately subgroup breast cancer cell lines, we used available expression data from 51 breast cancer cell lines. Recent work by Neve *et al*. showed that the recurrent genomic and transcriptional characteristics of breast cancer cell lines mirror those of primary breast tumors [31]. These investigators performed Affymetrix gene expression profiling on a set of 51 ER-positive and ER-negative breast cancer cell lines and used hierarchical clustering to show that the cell lines clustered into three main groups: basal A, basal B, and luminal [31]. We used this expression information from breast cancer cell lines to determine whether our list of 52 kinases would group these cell lines into the similar groups as with the profiling of human tumors (cell cycle checkpoint control, S6 kinase, immunomodulatory, and MAPKs groups). When hierarchical cluster analysis was performed on the expression data from these 51 cell lines using the list of 52 kinases identified here, the cell lines were accurately clustered into ER-positive or ER-negative groups (**Figure 3.5A**). Furthermore, the overexpressed kinases were able to subset ER-negative breast cancer cell lines into 4 groups in an unsupervised manner (**Figure 3.5B)**. These results indicate that the expression profile of the identified kinases is sufficiently robust to accurately discriminate between most ER-positive and ER-negative breast cancer cell lines and may serve as a reliable diagnostic tool to categorize human tumors in the future.

Figure 3.5- List of kinases validates in an independent data set of human breast cancer cell lines. (**A***)* Publically available breast cancer cell line expression data was clustered in an unsupervised manner using only the 52 kinase genes identified in our analysis. Unsupervised hierarchical clustering using the 52 kinases identified as being at least 2 fold more highly expressed clusters ER-positive from ER-negative breast cancer cell lines and identifies the luminal, basal A, and basal B subtypes. (**B**) Clustering only of ER-negative breast cancer cell lines using the 52 kinases identifies 4 subsets of ER-negative breast cancer cell lines. Analysis done by Corey Speers with input from Dr. Anna Tsimelzon and Dr. Susan Hilsenbeck.

Figure 3.5


A
AU-565
ZR-75-B
UACC812
ZR75-1
BT474
MDA-MB361
HCC1428
BT483
SUM225
HCC1007
SUM-52PE
600MPE
MDA-MB175-V
ZR75-30
SUM-185PE
SUM-44PE
HCC202
CAMA-1
MDA-MB453
T47D
MDA-MB415
HCC2185
DU4475
SKBR3
LY2
MCF7
MDA-MB157
HBL100
BT549
MDA-MB436
HCC-38
HCC1187
HCC1569
HCC1143
HCC2157
HCC70
HCC1008
HCC1599
HCC1937
HCC3153
MDA MB-468
MCF-12A
MCF10A
HCC1500
SUM-159PT
MDA MB-231
BT20
SUM-190PT
SUM-149PT
MDA-MB435
SUM-1315-MO
HS578T
HCC1954
MDA-MB134-V
ER
PR
HER2
cluster
guanylate kinase 1 /// guanylate kinase
abl-interactor 1
death-associated protein kinase 1
chemokine (C-X-C motif) ligand 10
LIM domain kinase 2
serine/threonine kinase 38 like
pyridoxal (pyridoxine, vitamin B6) kina
EPH receptor B4
phosphatidylinositol 4-kinase, catalytic
v-raf-1 murine leukemia viral oncogene
phosphoribosyl pyrophosphate synthe
phosphofructokinase, platelet
tyrosine 3-monooxygenase/tryptophan
serine/threonine kinase 17b (apoptosis
interleukin-1 receptor-associated kinas
serum/glucocorticoid regulated kinase
selenophosphate synthetase 1
maternal embryonic leucine zipper kina
UDP-glucose pyrophosphorylase 2
vaccinia related kinase 2
protein kinase, X-linked /// protein kina
uridine-cytidine kinase 2
SFRS protein kinase 1
CHK1 checkpoint homolog (S. pombe)
TTK protein kinase
BUB1 budding uninhibited by benzimic
v-yes-1 Yamaguchi sarcoma viral relat
adenylate kinase 2
pim-1 oncogene /// pim-1 oncogene
v-yes-1 Yamaguchi sarcoma viral onc
PTK7 protein tyrosine kinase 7
general transcription factor IIH, polype
myelin protein zero-like 1
ret finger protein
mitogen-activated protein kinase 1
RYK receptor-like tyrosine kinase
phosphoglycerate kinase 1
mitogen-activated protein kinase kinas
PI-3-kinase-related kinase SMG-1
IL2-inducible T-cell kinase
chemokine (C-C motif) ligand 4
lymphocyte-specific protein tyrosine ki
-3.0 -2.7 -2.5 -2.2 -1.9 -1.6 -1.4 -1.1 -0.8 -0.5 -0.3 0 0.3 0.5 0.8 1.1 1.4 1.6 1.9 2.2 2.5 2.7 3.0
B
BT20
MDA MB-468
MDA MB-231
SUM-159PT
HCC1500
MCF-12A
MCF10A
SUM-1315-MO2
SUM-149PT
MDA-MB435
MDA-MB157
MDA-MB436
BT549
HCC1954
HS578T
HCC-38
HBL100
HCC2185
SUM-190PT
SUM225
HCC202
SUM-185PE
MDA-MB453
AU-565
SKBR3
HCC3153
HCC1937
HCC1143
HCC70
HCC1008
HCC1599
HCC1187
HCC1569
HCC2157
ER
PR
HER2
cluster
Group 1 (cell cycle checkpoint)
Group 2 (S6 kinase)
Group 3 (immunomodulatory)
Group 4 (MAPK)

3.3.9 Kinase Knockdown Exerts Differential Growth Effects on ER-negative and ER-positive Breast Cancer Cell Lines

While the expression array profiling data allowed us to evaluate which kinases were differentially expressed, we wanted to determine whether these kinases are critical for the growth of ER-negative breast cancer and thus potentially tractable targets for the treatment of ER-negative disease. To do this we performed siRNA knockdown studies to determine the effect of individual kinase knockdown on breast cancer cell proliferation. ER-positive (MCF-7 and T47D) and ER-negative (MDA-MB-468 and MDA-MB-231, both "triple-negative") cells were transfected with siRNAs for 20 of the 52 kinases identified in our screen. All siRNA constructs used in the study showed at least 70% knockdown of target kinase expression for 4 days after transfection (representative examples are shown in **Figure 3.6**).

Of the 20 kinases evaluated, 14 were critical for the growth of ER-negative breast cancer. Knockdown of 9 (*EPHB4, LIMK2, DAPK1, YES1, RYK, VRK2, PTK7, RAF1, UCK2)* had a significant growth-inhibitory effect on ER-negative breast cancer (MDA-MB-468 and MDA-MB-231) but had little or no effect on ER-positive breast cancer cells. An additional 5 of 20 kinases **(***BUB1, CHK1, IRAK1, CCL4, TTK***)** inhibited growth of all breast cancer cell lines. Knockdown of 5 of the 20 kinases **(***STK38L, PIM1, SFRS1, PKXL, TLR1***)** had no effect on any breast cancer cell line growth, while knockdown of 1 of 20 kinases (*MPZL1*) had a significant growth-stimulatory effect on all breast cancer cell lines

Figure 3.6

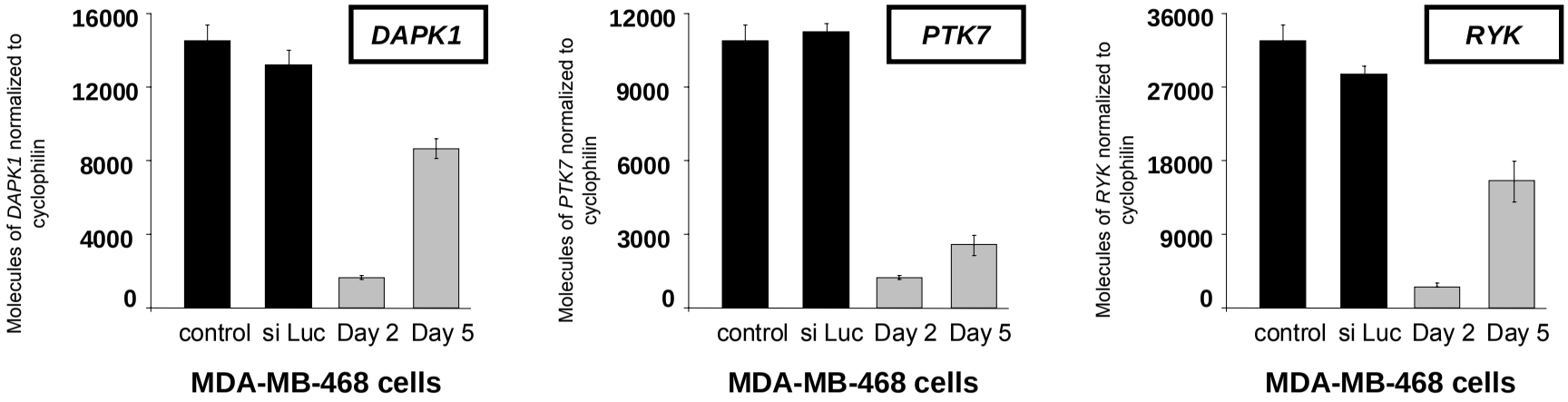


Figure 3.6 - siRNA constructs knockdown gene expression greater than 70% in all cell lines tested. Knockdown of target kinase expression was achieved using siRNA against the identified kinases, with representative data of *DAPK1*, *PTK7*, and *RYK* knockdown in MDA-MB-468 cells shown. Knockdown was confirmed by Q-PCR at day 2 and day 5 and was >70% in all experiments. Data are represented as mean ± SD.

examined. Representative growth curves from these knockdown experiments are shown in **Figure 3.7**.

Knockdown of many of the kinases in the "cell cycle checkpoint" cluster of ER-negative breast cancer had a profound inhibitory effect on ER-negative breast cancer cell growth but no effect on ER-positive breast cancer, while knockdown of certain kinases in the "immunomodulatory" cluster inhibited the growth of all breast cancer cell lines examined. A summary of results is shown in **Figure 3.8**, with bolded genes exhibiting a differential growth phenotype between ER-negative and ER-positive breast cancer cell lines. These results indicate that many of the kinases found to be highly expressed in ER-negative breast cancers are indeed critical for breast cancer cell growth.

3.3.10 S6 kinase Subtype of ER-negative Breast Cancer Predicts Poor Metastasis-Free Survival

To determine whether the identified list of differentially-expressed kinases provided prognostic information, we analyzed the survival data from the Wang and van de Vijver data sets using the kinases that we identified as being overexpressed in ER-negative breast tumors [3, 20]. The Wang dataset was obtained using breast cancer samples from patients with lymph-node negative breast cancer who were treated with breast conserving surgery or modified radical mastectomies from 1980–95. These patients also received radiotherapy when indicated, but did not receive systemic chemotherapy or hormonal adjuvant therapy. This time period was also prior to the

Figure 3.7- Effect of siRNA knockdown on the growth of ER-negative and ER-positive breast cancer cells. *DAPK1*, *PTK7*, and *RYK* knockdown inhibited growth in the ER-negative breast cancer cell lines MDA-MB-468 and MDA-MB-231 but not in the ER-positive breast cancer cell lines MCF-7 and T47D. Asterisk denotes significant difference in curves between kinase of interest knockdown and mock transfected growth curves, *P*-value < 0.05. Data are represented as mean ± SD.

Figure 3.7

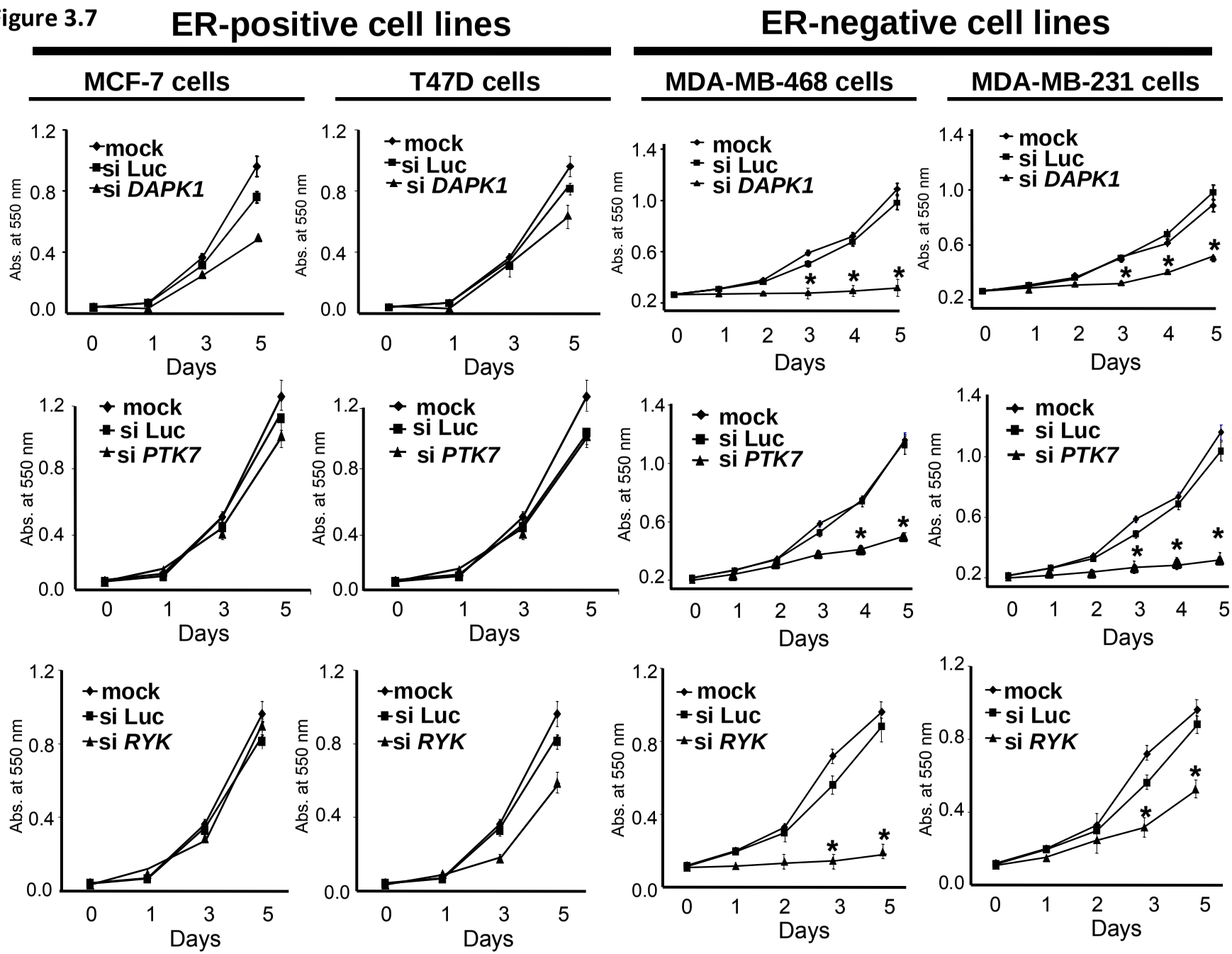

ER-positive cell lines
ER-negative cell lines
MCF-7 cells
T47D cells
MDA-MB-468 cells
MDA-MB-231 cells
mock
si Luc
si DAPK1
si PTK7
si RYK
Abs. at 550 nm
Days
*

Figure 3.8 - Effect of siRNA knockdown on the growth of ER-negative and ER-positive breast cancer cells. Similar growth inhibitory effects were seen with other kinases identified in the expression profiling. Effect of growth inhibition was based on the percentage of growth at day 4 compared to mock transfected controls. Cut-off percentages for inhibition, slight inhibition, no effect, and enhanced growth are shown. Bolded genes exhibited a differential growth phenotype between ER-negative and ER-positive breast cancer cell lines. Data are represented as mean ± SD.

Figure 3.8

	ER-positive cell lines		ER-negative cell lines	
	MCF-7	T47D	MDA-MB-231	MDA-MB-468
S6 kinase cluster (ER-negative, HER2-negative)				
EPHB4	**no effect**	**no effect**	**slight inhibit.**	**inhibit**
LIM domain kinase 2	**slight inhibit.**	**no effect**	**slight inhibit.**	**inhibit**
death-associated protein kinase 1	**slight inhibit.**	**slight inhibit.**	**inhibit**	**inhibit**
pim-1 oncogene	no effect	no effect	slight inhibit.	slight inhibit.
serine/threonine kinase 38 like (NDR2)	no effect	no effect	no effect	no effect
Cell Cycle checkpoint cluster (ER-negative, HER2-negative)				
maternal embryonic leucine zipper kinase	**slight inhibit.**	**slight inhibit.**	**inhibit**	**inhibit**
RYK receptor-like tyrosine kinase	**no effect**	**slight inhibit.**	**inhibit**	**inhibit**
uridine-cytidine kinase 2	**no effect**	**no effect**	**inhibit**	**inhibit**
vaccinia related kinase	**no effect**	**no effect**	**no effect**	**inhibit**
BUB1	inhibit	inhibit	inhibit	inhibit
CHEK1	inhibit	inhibit	inhibit	inhibit
adenylate kinase 2	inhibit	inhibit	inhibit	inhibit
tram trak kinase	inhibit	inhibit	inhibit	inhibit
selenophosphate synthetase 1	slight inhibit.	slight inhibit.	slight inhibit.	slight inhibit.
SFRS protein kinase 1	no effect	no effect	no effect	no effect
MAPK signaling cluster (ER-negative, HER2-positive)				
RAF-1	**no effect**	**slight inhibit.**	**inhibit**	**inhibit**
PTK7 protein tyrosine kinase 7	**no effect**	**no effect**	**inhibit**	**inhibit**
myelin protein zero-like 1	enhance growth	enhance growth	enhance growth	enhance growth
protein kinase, X linked	no effect	no effect	no effect	no effect
Immunomodulatory cluster (ER-negative, HER2-positive)				
yes-1	**slight inhibit.**	**slight inhibit.**	**inhibit**	**inhibit**
interleukin-1 receptor-associated kinase 1	inhibit	inhibit	inhibit	inhibit
chemokine (C-C motif) ligand 4	inhibit	inhibit	inhibit	inhibit
pyridoxal (pyridoxine, vitamin B6) kinase	no effect	slight inhibit.	inhibit	no effect
chemokine (C-X-C motif) ligand 10	no effect	no effect	no effect	slight inhibit.
toll-like receptor 1	no effect	no effect	no effect	no effect

Percent growth of control: **<50% inhibit** **50-85% slight inhibit.** **86-110% no effect** **>110% enhance** **ND: not done**

development of the anti-HER2 therapy, trastuzumab (Herceptin), and these patients were not treated with trastuzumab (Herceptin). 219 patients had undergone breast-conserving surgery and 67 modified radical mastectomy. Radiotherapy was given to 248 patients (87%), and metastasis free survival was tracked in all patients. In this data set, we first determined whether our list of 52 kinases overexpressed in ER-negative breast tumors could subcluster the Wang dataset tumors into the 4 subtypes of ER-negative tumors identified in our analysis. Hierarchical clustering of the ER-negative tumors from the Wang dataset using expression values of the kinases identified in this analysis did indeed identify 4 groups of ER-negative tumors (**Figure 3.9**). Figure of merit analysis showed that these four groups were stable against reclustering. Furthermore, these 4 clusters were similar in their kinase expression profiles to those identified previously, again identifying a S6 kinase signature cluster, a cell cycle checkpoint cluster, an immunomodulatory cluster, and a MAPK cluster. Kaplan-Meier analysis of the metastasis-free survival between the different subgroups of ER-negative tumors shows that women with the S6 kinase signature-expressing tumors have a much worse prognosis than the other groups, while women with breast cancers expressing the immunomodulatory kinases will have a much better prognosis (**Figure 3.10**). As a further means of validation, we performed the same analysis in the van de Vijver [3] data set and found similar results. In this dataset, all patients had stage I or II breast cancer and were younger than 53 years old; 151 had lymph-node–negative disease, and 144 had lymph node–positive disease. Ten of the 151 patients who had lymph-node–

Figure 3.9 - Hierarchical clustering and Kaplan-Meier metastasis free survival analysis of ER-negative tumors in the Wang *et al*. dataset. Hierarchical clustering of only ER-negative tumors identified the 4 clusters of ER-negative breast tumors in the Wang data set. The tumors were classified based on the expression level of the kinases identified in the analysis. Tumors that fell into the immunomodulatory cluster had a decreased risk of metastasis, and tumors in the cell cycle regulatory and S6 kinase clusters had a substantially elevated risk of metastasis at 5 years. Overall *P*-value was calculated based on the assumption that there would be no difference between any of the survival curves and was initially used to determine whether any one of the curves were significantly different. Further *P*-values were calculated comparing the designated two groups with the calculation of Chi square values. Immune refers to immunomodulatory group, CCC to the cell cycle checkpoint group, and S6 kinase to the S6 kinase group. Analysis done by Corey Speers with input from Dr. Susan Hilsenbeck.

Figure 3.9

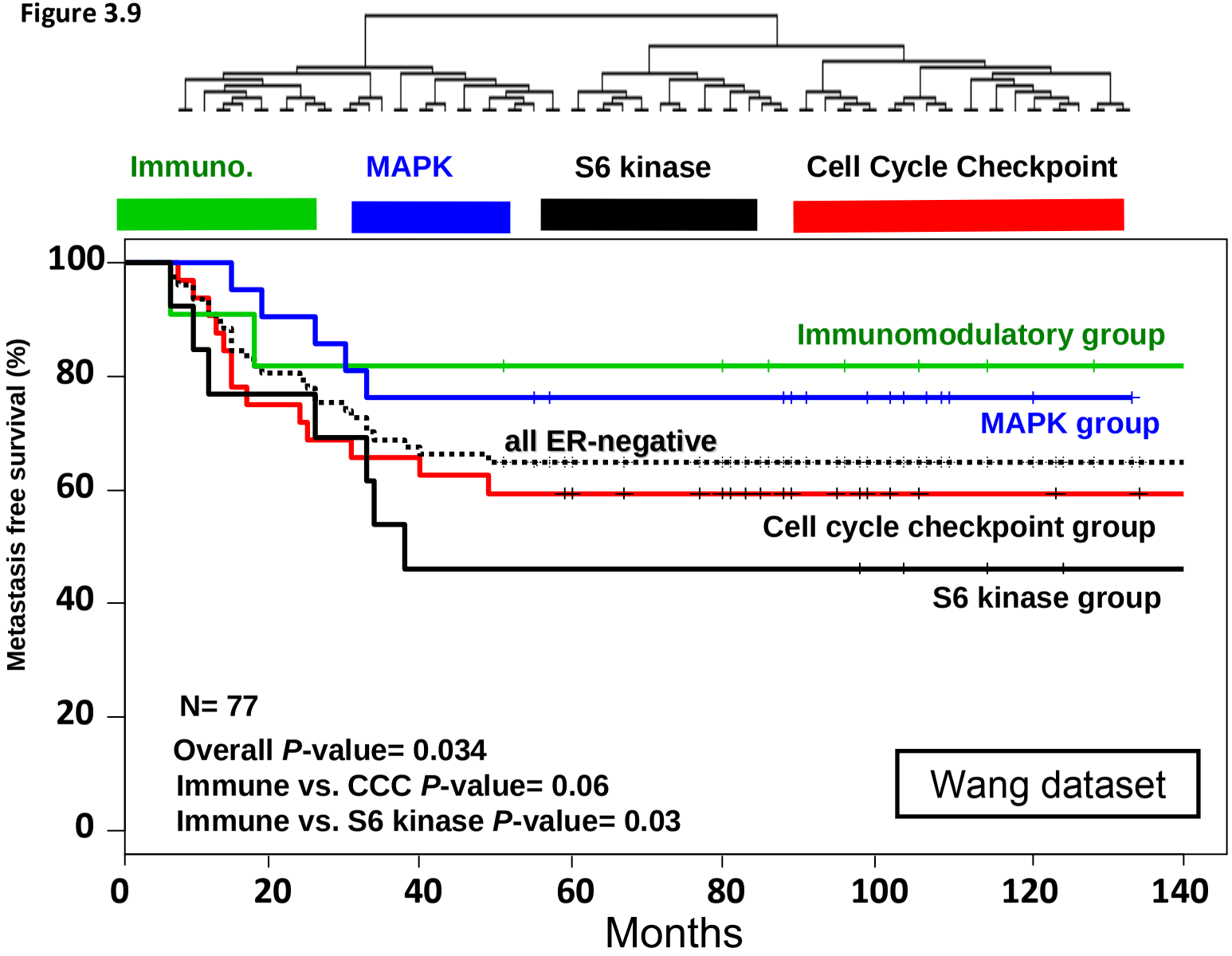
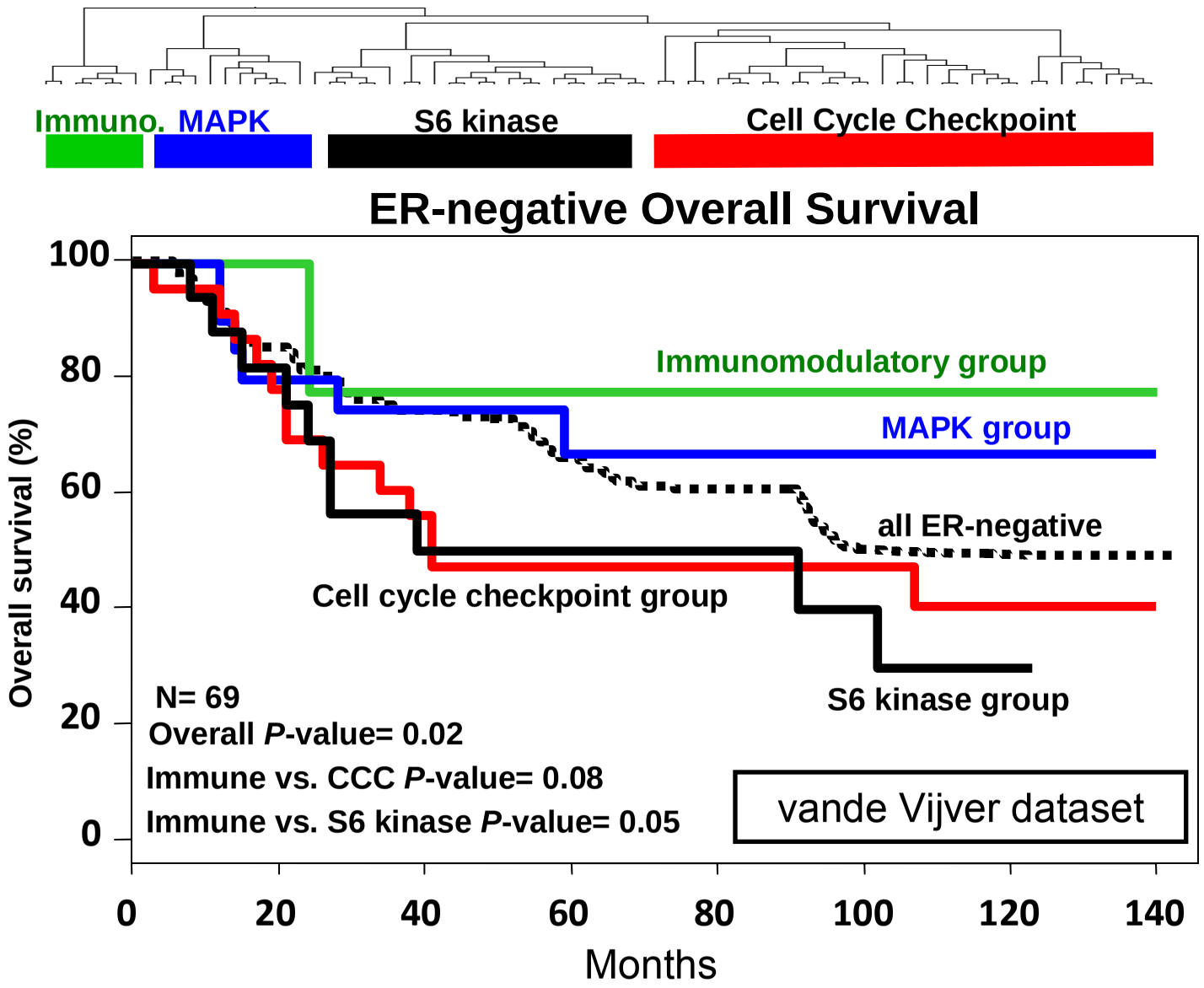

Immuno.
MAPK
S6 kinase
Cell Cycle Checkpoint
Metastasis free survival (%)
100
80
60
40
20
0
Immunomodulatory group
MAPK group
all ER-negative
Cell cycle checkpoint group
S6 kinase group
N= 77
Overall P-value= 0.034
Immune vs. CCC P-value= 0.06
Immune vs. S6 kinase P-value= 0.03
Wang dataset
0
20
40
60
80
100
120
140
Months

Figure 3.10 - Hierarchical clustering and Kaplan-Meier metastasis free survival analysis of ER-negative tumors in the van de Vijver *et al*. dataset. Similar results were found when hierarchical clustering was done in the van de Vijver data set. Overall survival was substantially higher in the immunomodulatory group than in the S6 kinase or cell cycle checkpoint groups. Overall *P*-value was calculated based on the assumption that there would be no difference between any of the survival curves and was initially used to determine whether any one of the curves were significantly different. Further *P*-values were calculated comparing the designated two groups with the calculation of Chi square values. Immune refers to immunomodulatory group, CCC to the cell cycle checkpoint group, and S6 kinase to the S6 kinase group. Analysis done by Corey Speers with input from Dr. Susan Hilsenbeck.

Figure 3.10


Immuno.
MAPK
S6 kinase
Cell Cycle Checkpoint
ER-negative Overall Survival
Overall survival (%)
100
80
60
40
20
0
Immunomodulatory group
MAPK group
all ER-negative
Cell cycle checkpoint group
S6 kinase group
N= 69
Overall P-value= 0.02
Immune vs. CCC P-value= 0.08
Immune vs. S6 kinase P-value= 0.05
vande Vijver dataset
0
20
40
60
80
100
120
140
Months

negative disease and 120 of the 144 who had lymph-node–positive disease had received adjuvant systemic therapy consisting of chemotherapy (90 patients), hormonal therapy (20), or both (20). As with the Wang dataset, hierarchical clustering of the ER-negative tumors identified the four groups of ER-negative tumors, which were again stable against reclustering, and, as with the Wang dataset, patients whose tumors had high expression of the immunomodulatory kinases had a significantly better overall survival than those with high expression of the S6 kinase and cell cycle checkpoint clusters. These data suggest that our list of differentially-expressed kinases can be used to identify distinct subtypes of ER-negative breast tumors, and that the tumor clusters defined by the expression of these kinases have either a good prognosis (immunomodulatory group) or a particularly poor prognosis (S6 kinase signature group) based on their kinase expression profile.

3.4 Discussion

In this report we show that Affymetrix gene expression profiling of human breast tumors is able to identify kinases that are differentially-expressed in ER-negative breast cancers as compared to ER-positive breast cancers. Further analysis revealed that ER-negative tumors can be clustered into 4 distinct groups, depending on the specific kinases expressed and the level of their expression. Analysis of publicly available breast tumor data sets confirmed that these kinases are indeed upregulated in ER-negative breast cancer. Studies in which knockdown of selected kinases using siRNA were conducted and demonstrated that many of the identified kinases are critical for ER-negative, including "triple-negative", breast cancer growth. Finally, analysis of kinase expression in human breast tumors demonstrated that the individual subtypes of ER-negative breast cancer identified by their kinase profile here have different outcomes. Specifically, these studies demonstrate that women whose ER-negative tumors highly express the kinases from the S6 kinase group have a particularly bad prognosis, while women whose tumors highly express immunomodulatory kinases have a relatively good prognosis. Such results suggest that characterization of human tumors based on kinase expression can be used to select patients appropriate for novel therapies. In addition, this study identifies potential targets for the treatment of ER-negative breast cancer, including the aggressive "triple-negative" form of breast cancer.

This is the first report to show that ER-negative breast cancers can be subdivided into biologically distinct groups based on expression levels of specific kinases. Our data

indicate that ER-negative breast tumors can be subdivided into 4 distinct groups, of particular importance are group 2 (S6 kinase group) and group 3 (immunomodulatory group), and that patients whose tumors express these kinases have very different prognoses. The immunomodulatory group (group 3) identified in this report has recently become of a focus of increasing scientific inquiry. In this report, we show differential expression of these immunomodulatory kinases in the epithelial compartment (as demonstrated by high expression in breast cancer cell lines grown *in vitro*). There remains a question of whether these kinases are also expressed in the non-epithelial cells present in breast tumors, specifically in infiltrating immune cells. Two lines of evidence suggest that this difference is predominantly from the epithelial compartment. First, recent work by Neve *et al*. [31] validates the differential expression identified in this report in ER-negative breast cancer epithelial cell lines as compared to their ER-positive cell line counterparts. Their experiments were conducted using a purified, homogenous population of breast cancer epithelial cell lines that show the same differential expression we note in our human tumor studies. Secondly, the siRNA knockdown experiments reported herein also show that knockdown of these immunomodulatory kinases *in vitro* in epithelial breast cancer cell lines have a differential effect on cell growth. We are currently examining whether there are differences in the immune and stomal components of the different subtypes of ER-negative breast cancer, including the immunomodulatory subtype, identified in this study.

The role of the immune system in cancer has historically investigated how the immune system itself responds to the "foreign" cancer as the primary focus. It is now being appreciated that the tumor itself may act autonomously to influence the stromal microenvironment and evade recognition by the immunosurveillance machinery. It is possible that the immune-regulatory genes expressed by the epithelial cancer cells affect this local immune response to these tumors. Recent work by Teschendorff *et al.* supports our findings [32]. This group also identified an immunomodulatory profile in ER-negative breast cancer which was shown to confer better prognosis [32]. It will be interesting to investigate in the future whether modulation of intrinsic gene expression by the tumor is an important mechanism by which cancer cells can avoid immunosurveillance, including the proper controls meant to keep aberrant growth in check [32, 33].

These studies provide a large number of promising new targets for the treatment of ER-negative breast cancer. ER-positive breast cancers are now routinely treated using SERMs and aromatase inhibitors, and these cancers are now even prevented using such pharmacologic intervention [9]. Recent studies have shown that intrinsic breast cancer subtypes differ depending on the ethnicity of the patient from whom the tumor is obtained. Carey *et al.* refined an IHC-based assay to categorize the prevalence of varying breast cancer subtypes in different populations [8]. It was shown that the prevalence of the basal-like subtypes was strongly influenced by race and menopause status. The highest prevalence of basal-like tumors was noted in

premenopausal African American breast cancer patients [8, 34]. Basal-like tumors, which are almost uniformly ER-negative, PR-negative, and HER2 negative ("triple-negative"), are more aggressive, carry a higher proliferative capacity, occur at a younger age, and carry a particularly bad prognosis [34, 35]. This work provides the rationale for targeted therapy using multi-kinase inhibitors to treat this type of breast cancer more prevalent among a traditionally underserved population.

One particularly promising agent for the treatment of triple-negative breast cancer is the multiple kinase inhibitor dasatinib. Dasatinib is an oral kinase inhibitor that inhibits several kinases, including c-Src, BCR/Abl, *YES1*, and *EPHB4*, and is currently approved for the treatment of chronic myelogenous leukemia (CML) and acute lymphoblastic leukemia (ALL). Intriguingly, this kinase inhibitor blocks the activity of many of the kinases identified in this analysis including c-Src, BCR/Abl, *YES1*, *EPHB4*, as well as *KIT,* and *EPHA2*. There is increasing pre-clinical and clinical data to suggest that this multi-kinase inhibitor may be an effective treatment for triple-negative breast cancer. Initial experiments in both prostate and breast cancer cell lines demonstrated that dasatinib significantly inhibited breast cancer cell line growth [36]. Further i*n vitro* experimentation shows that dasatinib is especially efficacious at inhibiting basal-like and post-EMT ER-negative breast cancer cell line growth and these studies led to the identification potential biomarkers of response [37]. Clinical trials are currently being conducted, with preliminary promising results already being presented, using dasatinib in women with ER-negative breast cancer in the metastatic setting.

The results reported here demonstrate that genomic profiling of human breast cancers can identify subtypes of ER-negative breast cancer, but even more importantly, can also identify new targets for effective treatment of these aggressive breast cancers. Given the current difficulty in treating ER-negative breast cancer, and particularly the triple-negative form of breast cancer, the identification of the kinases that are critical for the growth of these cancers represents the first step towards effective individualized targeted therapy for these poor prognosis ER-negative breast cancers.

References

1. Chang, J.C., et al., *Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer.* Lancet, 2003. **362**(9381): p. 362-9.
2. Chang, J.C., et al., *Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients.* J Clin Oncol, 2005. **23**(6): p. 1169-77.
3. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.
4. van 't Veer, L.J., et al., *Expression profiling predicts outcome in breast cancer.* Breast Cancer Res, 2003. **5**(1): p. 57-8.
5. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-6.
6. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
7. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.
8. Carey, L.A., et al., *Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.* Jama, 2006. **295**(21): p. 2492-502.
9. Swaby, R.F., C.G. Sharma, and V.C. Jordan, *SERMs for the treatment and prevention of breast cancer.* Rev Endocr Metab Disord, 2007. **8**(3): p. 229-39.
10. Tan, A.R. and S.M. Swain, *Adjuvant chemotherapy for breast cancer: an update.* Semin Oncol, 2001. **28**(4): p. 359-76.
11. Miller, K., et al., *Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer.* N Engl J Med, 2007. **357**(26): p. 2666-76.
12. Viani, G.A., et al., *Adjuvant trastuzumab in the treatment of her-2-positive early breast cancer: a meta-analysis of published randomized trials.* BMC Cancer, 2007. **7**: p. 153.
13. Dowell, J.E. and J.D. Minna, *EGFR mutations and molecularly targeted therapy: a new era in the treatment of lung cancer.* Nat Clin Pract Oncol, 2006. **3**(4): p. 170-1.
14. Herbst, R.S., M. Fukuoka, and J. Baselga, *Gefitinib--a novel targeted approach to treating cancer.* Nat Rev Cancer, 2004. **4**(12): p. 956-65.
15. Minna, J.D. and J. Dowell, *Erlotinib hydrochloride.* Nat Rev Drug Discov, 2005. **Suppl**: p. S14-5.
16. Dancey, J. and E.A. Sausville, *Issues and progress with protein kinase inhibitors for cancer treatment.* Nat Rev Drug Discov, 2003. **2**(4): p. 296-313.
17. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.

18. Chan, E.Y., S. Kir, and S.A. Tooze, *siRNA screening of the kinome identifies ULK1 as a multidomain modulator of autophagy.* J Biol Chem, 2007. **282**(35): p. 25464-74.
19. Draviam, V.M., et al., *A functional genomic screen identifies a role for TAO1 kinase in spindle-checkpoint signalling.* Nat Cell Biol, 2007. **9**(5): p. 556-64.
20. Wang, Y., et al., *Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.* Lancet, 2005. **365**(9460): p. 671-9.
21. Miller, L.D., et al., *An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival.* Proc Natl Acad Sci U S A, 2005. **102**(38): p. 13550-5.
22. Hess, K.R., et al., *Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer.* J Clin Oncol, 2006. **24**(26): p. 4236-44.
23. Yu, K., et al., *A modular analysis of breast cancer reveals a novel low-grade molecular signature in estrogen receptor-positive tumors.* Clin Cancer Res, 2006. **12**(11 Pt 1): p. 3288-96.
24. Richardson, A.L., et al., *X chromosomal abnormalities in basal-like human breast cancer.* Cancer Cell, 2006. **9**(2): p. 121-32.
25. Chin, K., et al., *Genomic and transcriptional aberrations linked to breast cancer pathophysiologies.* Cancer Cell, 2006. **10**(6): p. 529-41.
26. Ivshina, A.V., et al., *Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer.* Cancer Res, 2006. **66**(21): p. 10292-301.
27. Ginestier, C., et al., *Prognosis and gene expression profiling of 20q13-amplified breast cancers.* Clin Cancer Res, 2006. **12**(15): p. 4533-44.
28. Minn, A.J., et al., *Genes that mediate breast cancer metastasis to lung.* Nature, 2005. **436**(7050): p. 518-24.
29. Bild, A.H., et al., *Oncogenic pathway signatures in human cancers as a guide to targeted therapies.* Nature, 2006. **439**(7074): p. 353-7.
30. Whitlock, M.C., *Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach.* J Evol Biol, 2005. **18**(5): p. 1368-73.
31. Neve, R.M., et al., *A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes.* Cancer Cell, 2006. **10**(6): p. 515-27.
32. Teschendorff, A.E., et al., *An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer.* Genome Biol, 2007. **8**(8): p. R157.
33. Amini, R.M., et al., *Mast cells and eosinophils in invasive breast carcinoma.* BMC Cancer, 2007. **7**(1): p. 165.
34. Furberg, H., et al., *Tumor characteristics in African American and white women.* Breast Cancer Res Treat, 2001. **68**(1): p. 33-43.
35. Porter, P.L., et al., *Racial differences in the expression of cell cycle-regulatory proteins in breast carcinoma.* Cancer, 2004. **100**(12): p. 2533-42.

36. Nam, S., et al., *Action of the Src family kinase inhibitor, dasatinib (BMS-354825), on human prostate cancer cells.* Cancer Res, 2005. **65**(20): p. 9185-9.
37. Finn, R.S., et al., *Dasatinib, an orally active small molecule inhibitor of both the src and abl kinases, selectively inhibits growth of basal-type/"triple-negative" breast cancer cell lines growing in vitro.* Breast Cancer Res Treat, 2007. **105**(3): p. 319-26.

Supplementary table 3.1- All kinases and associated genes used in the expression profiling

GeneSymbol	Accession	LocusLink
AAK1	NM_014911	22848
AATK	XM_290778	9625
ABL1	NM_005157	25
ABL2	NM_005158	27
ACK1	NM_005781	10188
ACVR1	NM_001105	90
ACVR1B	NM_004302	91
ACVR2	NM_001616	92
ACVR2B	NM_001106	93
ACVRL1	NM_000020	94
ADAM9	NM_003816	8754
ADCK1	NM_020421	57143
ADCK2	NM_052853	90956
ADCK2	NM_024876	79934
ADCK5	NM_174922	203054
ADK	NM_001123	132
ADRA1A	NM_000680	148
ADRA1B	NM_000679	147
ADRB2	NM_000024	154
ADRBK1	NM_001619	156
ADRBK2	NM_005160	157
AGTR2	NM_000686	186
AK1	NM_000476	203
AK2	NM_001625	204
AK3	NM_013410	205
AK3L1	NM_016282	50808
AK5	NM_012093	26289
AK7	NM_152327	122481
AKAP1	NM_003488	8165
AKAP11	NM_016248	11215
AKAP13	NM_006738	11214
AKAP3	NM_006422	10566
AKAP4	NM_003886	8852
AKAP5	NM_004857	9495
AKAP6	NM_004274	9472
AKAP7	NM_004842	9465
AKAP8	NM_005858	10270
AKT1	NM_005163	207
AKT2	NM_001626	208
AKT3	NM_005465	10000
ALK	NM_004304	238
ALK7	NM_145259	130399
ALS2CR2	NM_018571	55437
ALS2CR7	NM_139158	65061
AMHR2	NM_020547	269
ANGPT4	NM_015985	51378
ANKK1	NM_178510	255239
ANKRD3	NM_020639	54101
APEG1	NM_005876	57518
APPL	NM_012096	26060
ARAF1	NM_001654	369
ARK5	NM_014840	9891
ASK	NM_006716	10926
ASP	NM_031916	83853
ATM	NM_000051	472
ATR	NM_001184	545
AURKB	NM_004217	9212
AURKC	NM_003160	6795
AVPR1A	NM_000706	552
AVPR1B	NM_000707	553
AXL	NM_001699	558
AZU1	NM_001700	566
BCKDK	NM_005881	10295
BCR	NM_004327	613
BDKRB2	NM_000623	624
BLK	NM_001715	640
BLNK	NM_013314	29760
BMP2K	NM_017593	55589
BMPR1A	NM_004329	657
BMPR1B	NM_001203	658
BMPR2	NM_001204	659
BMX	NM_001721	660
BRAF	NM_004333	673
BRD2	NM_005104	6046
BRDT	NM_001726	676
BTK	NM_000061	695
BUB1	NM_004336	699
BUB1B	NM_001211	701
C14ORF20	NM_174944	283629
C20ORF64	NM_033550	112858
C20ORF97	NM_021158	57761
C6ORF199	NM_145025	221264
C7ORF16	NM_006658	10842

C8FW	NM_025195	10221
C9ORF12	NM_022755	64768
CALM3	NM_005184	5509
CAMK1	NM_003656	8536
CAMK1D	NM_020397	57118
CAMK1G	NM_020439	57172
CAMK2A	NM_015981	815
CAMK2B	NM_001220	816
CAMK2D	NM_001221	817
CAMK2G	NM_001222	818
CAMK4	NM_001744	814
CAMKK1	NM_172206	84254
CAMKK2	NM_006549	10645
CARD10	NM_014550	29775
CARD14	NM_024110	79092
CARK	NM_015978	51086
CARKL	NM_013276	23729
CASK	NM_003688	8573
CCL2	NM_002982	6347
CCL4	NM_002984	6351
CCRK	NM_012119	23552
CD3E	NM_000733	916
CD4	NM_000616	920
CD7	NM_006137	924
CDACD1	NM_030911	81602
CDC2	NM_001786	983
CDC2L1	NM_001787	984
CDC2L2	NM_024011	985
CDC2L5	NM_003718	8621
CDC42BPA	NM_003607	8476
CDC42BPB	NM_006035	9578
CDC7L1	NM_003503	8317
CDK10	NM_003674	8558
CDK11	NM_015076	23097
CDK2	NM_001798	1017
CDK3	NM_001258	1018
CDK4	NM_000075	1019
CDK5	NM_004935	1020
CDK5R1	NM_003885	8851
CDK5R2	NM_003936	8941
CDK5RAP1	NM_016082	51654
CDK5RAP3	NM_025197	80279

CDK6	NM_001259	1021
CDK7	NM_001799	1022
CDK8	NM_001260	1024
CDK9	NM_001261	1025
CDKL1	NM_004196	8814
CDKL2	NM_003948	8999
CDKL3	NM_016508	51265
CDKL5	NM_003159	6792
CDKN1A	NM_000389	1026
CDKN1B	NM_004064	1027
CDKN1C	NM_000076	1028
CDKN2B	NM_004936	1030
CDKN2C	NM_001262	1031
CDKN2D	NM_001800	1032
CDKN3	NM_005192	1033
CERK	NM_022766	64781
CGEF2	NM_007023	11069
CHEK1	NM_001274	1111
CHEK2	NM_007194	11200
CHK	NM_001277	1119
CHKL	NM_005198	1120
CHRM1	NM_000738	1128
CHUK	NM_001278	1147
CINP	NM_032630	51550
CIT	NM_007174	11113
CKB	NM_001823	1152
CKM	NM_001824	1158
CKMT1	NM_020990	1159
CKMT2	NM_001825	1160
CKS1B	NM_001826	1163
CKS2	NM_001827	1164
CLK	NM_009905	12747
CLK1	NM_004071	1195
CLK2	NM_001291	1196
CLK3	NM_001292	1198
CLK4	NM_020666	57396
CNK1	NM_006314	10256
COL4A3BP	NM_005713	10087
COPB2	NM_004766	9276
CRK7	NM_016507	51755
CRKL	NM_005207	1399
CSF1R	NM_005211	1436

CSK	NM_004383	1445
CSNK1A1	NM_001892	1452
CSNK1D	NM_001893	1453
CSNK1E	NM_001894	1454
CSNK1G1	NM_022048	53944
CSNK1G2	NM_001319	1455
CSNK1G3	NM_004384	1456
CSNK2A1	NM_001895	1457
CSNK2A2	NM_001896	1459
CSNK2B	NM_001320	1460
CXCL10	NM_001565	3627
DAPK1	NM_004938	1612
DAPK2	NM_014326	23604
DAPK3	NM_001348	1613
DCAMKL1	NM_004734	9201
DCK	NM_000788	1633
DDR1	NM_001954	780
DDR2	NM_006182	4921
DGKA	NM_001345	1606
DGKB	NM_004080	1607
DGKD	NM_003648	8527
DGKE	NM_003647	8526
DGKG	NM_001346	1608
DGKI	NM_004717	9162
DGKQ	NM_001347	1609
DGKZ	NM_003646	8525
DGUOK	NM_001929	1716
DKFZP434C131	XM_044630	25989
DKFZP434C1418	NM_173655	285220
DKFZP586B1621	NM_015533	26007
DKFZP761P0423	XM_291277	157285
DKFZp761P1010	NM_018423	55359
DLG1	NM_004087	1739
DLG2	NM_001364	1740
DLG3	NM_021120	1741
DLG4	NM_001365	1742
DMPK	NM_004409	1760

DNAJC3	NM_006260	5611
DOK1	NM_001381	1796
DTYMK	NM_012145	1841
DUSP1	NM_004417	1843
DUSP10	NM_144728	11221
DUSP2	NM_004418	1844
DUSP22	NM_020185	56940
DUSP4	NM_057158	1846
DUSP5	NM_004419	1847
DUSP6	NM_001946	1848
DUSP7	NM_001947	1849
DUSP8	NM_004420	1850
DYRK1A	NM_001396	1859
DYRK1B	NM_004714	9149
DYRK2	NM_006482	8445
DYRK3	NM_003582	8444
DYRK4	NM_003845	8798
EDN2	NM_001956	1907
EEF2K	NM_013302	29904
EGFR	NM_005228	1956
EIF2AK3	NM_004836	9451
EIF2AK4	XM_031612	27104
EKI1	NM_018638	55500
EPAC	NM_006105	10411
EPHA1	NM_005232	2041
EPHA2	NM_004431	1969
EPHA3	NM_005233	2042
EPHA4	NM_004438	2043
EPHA5	NM_004439	2044
EPHA7	NM_004440	2045
EPHA8	NM_020526	2046
EPHB1	NM_004441	2047
EPHB2	NM_004442	2048
EPHB3	NM_004443	2049
EPHB4	NM_004444	2050
EPHB6	NM_004445	2051
ERBB2	NM_004448	2064
ERBB3	NM_001982	2065
ERBB4	NM_005235	2066
ERK8	NM_139021	225689
ERN1	NM_001433	2081
EVI1	NM_005241	2122

FASTK	NM_006712	10922
FER	NM_005246	2241
FES	NM_002005	2242
FGFR1	NM_000604	2260
FGFR2	NM_000141	2263
FGFR3	NM_000142	2261
FGFR4	NM_002011	2264
FGR	NM_005248	2268
FLJ10074	NM_017988	55681
FLJ10761	NM_018208	55224
FLJ10842	NM_018238	55750
FLJ11149	NM_018339	55312
FLJ12476	NM_022784	64799
FLJ13052	NM_023018	65220
FLJ14813	NM_032844	84930
FLJ20574	NM_017886	54986
FLJ22002	NM_024838	79896
FLJ23074	NM_025052	80122
FLJ23119	NM_024652	79705
FLJ23356	NM_032237	84197
FLJ25006	NM_144610	124923
FLJ32685	NM_152534	152110
FLJ32704	NM_152572	158067
FLJ34389	NM_152649	197259
FLJ35107	NM_182629	348825
FLT1	NM_002019	2321
FLT3	NM_004119	2322
FLT4	NM_002020	2324
FN3K	NM_022158	64122
FN3KRP	NM_024619	79672
FRAP1	NM_004958	2475
FRDA	NM_000144	2395
FRK	NM_002031	2444
FUK	NM_145059	197258
FYB	NM_001465	2533
FYN	NM_002037	2534
GAK	NM_005255	2580
GALK1	NM_000154	2584
GALK2	NM_002044	2585
GAP43	NM_002045	2596
GCK	NM_000162	2645
GFRA2	NM_001495	2675

GK	NM_000167	2710
GK2	NM_033214	2712
GMFB	NM_004124	2764
GMFG	NM_004877	9535
GNE	NM_005476	10020
GPRK2L	NM_005307	2868
GPRK5	NM_005308	2869
GPRK6	NM_002082	2870
GPRK7	NM_139209	131890
GSG2	NM_031965	83903
GSK3A	NM_019884	2931
GSK3B	NM_002093	2932
GTF2H1	NM_005316	2965
GUCY2C	NM_004963	2984
GUCY2D	NM_000180	3000
GUCY2F	NM_001522	2986
GUK1	NM_000858	2987
H11	NM_014365	26353
HAK	NM_052947	115701
HCK	NM_002110	3055
HIPK1	NM_152696	204851
HIPK2	NM_022740	28996
HIPK3	NM_005734	10114
HIPK4	NM_144685	147746
HK1	NM_000188	3098
HK2	NM_000189	3099
HK3	NM_002115	3101
HRI	NM_014413	27102
HSA250839	NM_018401	55351
HSMDPKIN	XM_290516	55561
HUNK	NM_014586	30811
ICAP-1A	NM_004763	9270
ICK	NM_014920	22858
IGF1R	NM_000875	3480
IHPK1	NM_153273	9807
IHPK2	NM_016291	51447
IHPK3	NM_054111	117283
IKBKAP	NM_003640	8518
IKBKB	XM_032491	3551
IKBKE	NM_014002	9641
IL2	NM_000586	3558
ILK	NM_004517	3611

ILKAP	NM_030768	80895
IMPK	NM_152230	253430
INSR	NM_000208	3643
INSRR	NM_014215	3645
IPMK	NM_152230	253430
IRAK1	NM_001569	3654
IRAK2	NM_001570	3656
IRAK3	NM_007199	11213
IRS1	NM_005544	3667
ITK	NM_005546	3702
ITPK1	NM_014216	3705
ITPKA	NM_002220	3706
ITPKB	NM_002221	3707
ITPKC	NM_025194	80271
JAK1	NM_002227	3716
JAK2	NM_004972	3717
JAK3	NM_000215	3718
JIK	NM_016281	51347
KDR	NM_002253	3791
KHK	NM_000221	3795
KIAA0551	XM_039796	23043
KIAA0561	XM_038150	23031
KIAA0999	NM_025164	23387
KIAA1361	XM_290796	57551
KIAA1399	XM_046685	57574
KIAA1639	XM_290923	57729
KIAA1765	XM_047355	85443
KIAA1804	NM_032435	84451
KIAA1811	NM_032430	84446
KIAA1883	XM_055866	114783
KIF13B	NM_015254	23303
KIS	NM_144624	127933
KIT	NM_000222	3815
KPI2	NM_014916	22853
KSR2	NM_173598	283455
LAK	NM_025144	80216
LATS1	NM_004690	9113
LATS2	NM_014572	26524
LCK	NM_005356	3932
LCP2	NM_005565	3937
LIM	NM_006457	10611
LIMK1	NM_002314	3984

LIMK2	NM_005569	3985
LOC115704	NM_145245	115704
LOC149420	NM_152835	149420
LOC221823	NM_175886	221823
LOC340371	NM_178564	340371
LOC91807	NM_182493	91807
LTK	NM_002344	4058
LYK5	NM_153335	92335
LYN	NM_002350	4067
MADH7	NM_005904	4092
MAGI-3	NM_020965	260425
MAK	NM_005906	4117
MALT1	NM_006785	10892
MAP2K1	NM_002755	5604
MAP2K1IP1	NM_021970	8649
MAP2K2	NM_030662	5605
MAP2K3	NM_002756	5606
MAP2K4	NM_003010	6416
MAP2K5	NM_002757	5607
MAP2K6	NM_002758	5608
MAP2K7	NM_145185	5609
MAP3K1	XM_042066	4214
MAP3K10	NM_002446	4294
MAP3K11	NM_002419	4296
MAP3K12	NM_006301	7786
MAP3K13	NM_004721	9175
MAP3K14	NM_003954	9020
MAP3K2	NM_006609	10746
MAP3K3	NM_002401	4215
MAP3K4	NM_005922	4216
MAP3K5	NM_005923	4217
MAP3K6	NM_004672	9064
MAP3K7	NM_003188	6885
MAP3K8	NM_005204	1326
MAP3K9	XM_027237	4293
MAP4K1	NM_007181	11184
MAP4K2	NM_004579	5871
MAP4K3	NM_003618	8491
MAP4K4	NM_004834	9448
MAP4K5	NM_006575	11183
MAPK1	NM_002745	5594
MAPK10	NM_002753	5602

MAPK11	NM_002751	5600
MAPK12	NM_002969	6300
MAPK13	NM_002754	5603
MAPK14	NM_001315	1432
MAPK3	NM_002746	5595
MAPK4	NM_002747	5596
MAPK6	NM_002748	5597
MAPK7	NM_002749	5598
MAPK8	NM_002750	5599
MAPK8IP1	NM_005456	9479
MAPK8IP2	NM_012324	23542
MAPK8IP3	NM_015133	23162
MAPK9	NM_002752	5601
MAPKAPK2	NM_004759	9261
MAPKAPK3	NM_004635	7867
MAPKAPK5	NM_003668	8550
MARK1	NM_018650	4139
MARK2	NM_004954	2011
MARK3	NM_002376	4140
MARKL1	NM_031417	57787
MAST205	NM_015112	23139
MATK	NM_002378	4145
MBIP	NM_016586	51562
MELK	NM_014791	9833
MERTK	NM_006343	10461
MET	NM_000245	4233
MGC16169	NM_033115	93627
MGC22688	NM_145001	202374
MGC26597	NM_152700	206426
MGC33182	NM_145203	122011
MGC42105	NM_153361	167359
MGC43306	XM_291304	169436
MGC45428	NM_152619	166614
MGC46424	NM_173492	138429
MGC4796	NM_032017	83931
MGC5601	NM_025247	80724
MGC8407	NM_024046	79012
MIDORI	NM_020778	57538
MINK	NM_015716	50488
MKNK1	NM_003684	8569
MKNK2	NM_017572	2872
MOS	NM_005372	4342

MPP1	NM_002436	4354
MPP2	NM_005374	4355
MPP3	NM_001932	4356
MPZL1	NM_003953	9019
MRC2	NM_006039	9902
MST1R	NM_002447	4486
MST4	NM_016542	51765
MUSK	NM_005592	4593
MVD	NM_002461	4597
MVK	NM_000431	4598
MYLK	NM_005965	4638
MYLK2	NM_033118	85366
MYO3A	NM_017433	53904
MYO3B	NM_138995	14069
NAGK	NM_017567	55577
NBEA	NM_015678	26960
NBP	NM_025233	80347
NEK1	XM_291107	4750
NEK11	NM_024800	79858
NEK2	NM_002497	4751
NEK3	NM_002498	4752
NEK4	NM_003157	6787
NEK6	NM_014397	10783
NEK7	NM_133494	140609
NEK8	NM_178170	284086
NEK9	NM_033116	91754
NLK	NM_016231	51701
NME1	NM_000269	4830
NME2	NM_002512	4831
NME3	NM_002513	4832
NME4	NM_005009	4833
NME5	NM_003551	8382
NME6	NM_005793	10201
NME7	NM_013330	29922
NPR1	NM_000906	4881
NPR2	NM_000907	4882
NRBP	NM_013392	29959
NRG3	XM_166086	10718
NTRK1	NM_002529	4914
NTRK2	NM_006180	4915
NTRK3	NM_002530	4916
NYD-SP25	NM_033516	89882

OSR1	NM_005109	9943
P15RS	NM_018170	55197
PACE-1	NM_020423	57147
PACSIN1	NM_020804	29993
PAG	NM_018440	55824
PAK1	NM_002576	5058
PAK2	NM_002577	5062
PAK3	NM_002578	5063
PAK4	NM_005884	10298
PAK6	NM_020168	56924
PAK7	NM_020341	57144
PANK1	NM_138316	53354
PANK3	NM_024594	79646
PANK4	NM_018216	55229
PAPSS1	NM_005443	9061
PAPSS2	NM_004670	9060
PASK	NM_015148	23178
PCK1	NM_002591	5105
PCK2	NM_004563	5106
PCTK1	NM_006201	5127
PCTK2	NM_002595	5128
PCTK3	NM_002596	5129
PDGFRA	NM_006206	5156
PDGFRB	NM_002609	5159
PDK1	NM_002610	5163
PDK2	NM_002611	5164
PDK3	NM_005391	5165
PDK4	NM_002612	5166
PDPK1	NM_002613	5170
PDXK	NM_003681	8566
PFKFB1	NM_002625	5207
PFKFB2	NM_006212	5208
PFKFB3	NM_004566	5209
PFKFB4	NM_004567	5210
PFKL	NM_002626	5211
PFKM	NM_000289	5213
PFKP	NM_002627	5214
PFTK1	NM_012395	5218
PGK1	NM_000291	5230
PGK2	NM_138733	5232
PHKA1	NM_002637	5255
PHKA2	NM_000292	5256

PHKG1	NM_006213	5260
PHKG2	NM_000294	5261
PI4K2B	NM_018323	55300
PI4KII	NM_018425	55361
PIK3C2A	NM_002645	5286
PIK3C2B	NM_002646	5287
PIK3C2G	NM_004570	5288
PIK3CA	NM_006218	5290
PIK3CB	NM_006219	5291
PIK3CG	NM_002649	5294
PIK3R2	NM_005027	5296
PIK3R3	NM_003629	8503
PIK3R4	NM_014602	30849
PIK4CA	NM_002650	5297
PIK4CB	NM_002651	5298
PIM1	NM_002648	5292
PIM2	NM_006875	11040
PINK1	NM_032409	65018
PIP5K1A	NM_003557	8394
PIP5K2A	NM_005028	5305
PIP5K2B	NM_003559	8396
PIP5K2C	NM_024779	79837
PITPNM3	NM_031220	83394
PKE	NM_173575	282974
PKIA	NM_006823	5569
PKIB	NM_032471	5570
PKLR	NM_000298	5313
PKM2	NM_002654	5315
PKMYT1	NM_004203	9088
PKNBETA	NM_013355	29941
PLK	NM_005030	5347
PMSCL2	NM_002685	5394
PMVK	NM_006556	10654
PNKP	NM_007254	11284
PPP1R1B	NM_032192	84152
PPP2CA	NM_002715	5515
PPP2CB	NM_004156	5516
PPP4C	NM_002720	5531
PRKAA1	NM_006251	5562
PRKAA2	NM_006252	5563
PRKACA	NM_002730	5566
PRKACB	NM_002731	5567

PRKACG	NM_002732	5568
PRKAG1	NM_002733	5571
PRKAG3	NM_017431	53632
PRKAR1A	NM_002734	5573
PRKAR2A	NM_004157	5576
PRKAR2B	NM_002736	5577
PRKCA	NM_002737	5578
PRKCABP	NM_012407	9463
PRKCB1	NM_002738	5579
PRKCD	NM_006254	5580
PRKCE	NM_005400	5581
PRKCG	NM_002739	5582
PRKCH	NM_006255	5583
PRKCI	NM_002740	5584
PRKCL1	NM_002741	5585
PRKCL2	NM_006256	5586
PRKCM	NM_002742	5587
PRKCN	NM_005813	23683
PRKCQ	NM_006257	5588
PRKCSH	NM_002743	5589
PRKCZ	NM_002744	5590
PRKD2	NM_016457	25865
PRKDC	NM_006904	5591
PRKG1	NM_006258	5592
PRKG2	NM_006259	5593
PRKR	NM_002759	5610
PRKRA	NM_003690	8575
PRKWNK1	NM_018979	65125
PRKWNK2	NM_006648	65268
PRKWNK3	NM_020922	65267
PRKWNK4	NM_032387	65266
PRKX	NM_005044	5613
PRKY	NM_002760	5616
PRPF4B	NM_003913	8899
PRPS1	NM_002764	5631
PRPS2	NM_002765	5634
PRPSAP1	NM_002766	5635
PRPSAP2	NM_002767	5636
PSK	NM_016151	51677
PSKH1	NM_006742	5681
PSKH2	NM_033126	85481
PTK2	NM_005607	5747

PTK2B	NM_004103	2185
PTK6	NM_005975	5753
PTK7	NM_002821	5754
PTK9	NM_002822	5756
PTK9L	NM_007284	11344
PTPN5	NM_032781	84867
PTPRG	NM_002841	5793
PTPRJ	NM_002843	5795
PTPRR	NM_002849	5801
PTPRT	NM_007050	11122
PXK	NM_017771	54899
PYCS	NM_002860	5832
RAC1	NM_006908	5879
RAF1	NM_002880	5894
RAGE	NM_014226	5891
RASGRF2	NM_006909	5924
RBSK	NM_022128	64080
RET	NM_000323	5979
RFP	NM_006510	5987
RHOK	NM_002929	6011
RIOK1	NM_031480	83732
RIOK3	NM_003831	8780
RIPK1	NM_003804	8737
RIPK2	NM_003821	8767
RIPK3	NM_006871	11035
RNASEL	NM_021133	6041
ROCK1	NM_005406	6093
ROCK2	NM_004850	9475
ROR1	NM_005012	4919
ROR2	NM_004560	4920
ROS1	NM_002944	6098
RP2	NM_006915	6102
RPS6KA1	NM_002953	6195
RPS6KA2	NM_021135	6196
RPS6KA3	NM_004586	6197
RPS6KA4	NM_003942	8986
RPS6KA5	NM_004755	9252
RPS6KA6	NM_014496	27330
RPS6KB1	NM_003161	6198
RPS6KB2	NM_003952	6199
RPS6KC1	NM_012424	26750
RPS6KL1	NM_031464	83694

RYK	NM_002958	6259
SAST	XM_032034	22983
SCAP1	NM_003726	8631
SCYL1	NM_020680	57410
SEPHS1	NM_012247	22929
SGK	NM_005627	6446
SGK2	NM_016276	10110
SGKL	NM_013257	23678
SHC1	NM_003029	6464
SIK2	NM_015191	23235
SLK	NM_014720	9748
SMG1	NM_014006	23049
SNARK	NM_030952	81788
SNF1LK	NM_173354	150094
SNK	NM_006622	10769
SNRK	NM_017719	54861
SOCS1	NM_003745	8651
SOCS5	NM_014011	9655
SPA17	NM_017425	53340
SPEC2	NM_020240	56990
SPHK1	NM_021972	8877
SPHK2	NM_020126	56848
SPS2	NM_012248	22928
SQSTM1	NM_003900	8878
SRC	NM_005417	6714
SRMS	NM_080823	6725
SRPK1	NM_003137	6732
SRPK2	NM_182691	6733
SSH3BP1	NM_005470	10006
SSTK	NM_032037	83983
STK10	NM_005990	6793
STK11	NM_000455	6794
STK16	NM_003691	8576
STK17A	NM_004760	9263
STK17B	NM_004226	9262
STK18	NM_014264	10733
STK19	NM_004197	8859
STK22B	NM_053006	23617
STK22C	NM_052841	81629
STK22D	NM_032028	83942
STK23	NM_014370	26576
STK24	NM_003576	8428

STK25	NM_006374	10494
STK29	NM_003957	9024
STK3	NM_006281	6788
STK31	NM_031414	56164
STK33	NM_030906	65975
STK35	NM_080836	140901
STK36	NM_015690	27148
STK38	NM_007271	11329
STK38L	NM_015000	23012
STK39	NM_013233	27347
STK4	NM_006282	6789
STK6	NM_003600	6790
SYK	NM_003177	6850
TAF1	NM_004606	6872
TAF1L	NM_153809	138474
TAO1	NM_004783	9344
TBK1	NM_013254	29110
TEC	NM_003215	7006
TEK	NM_000459	7010
TESK1	NM_006285	7016
TESK2	NM_007170	10420
TEX14	NM_031272	56155
TGFBR1	NM_004612	7046
TGFBR2	NM_003242	7048
TIE	NM_005424	7075
TJP2	NM_004817	9414
TK1	NM_003258	7083
TK2	NM_004614	7084
TLK1	NM_012290	9874
TLK2	NM_006852	11011
TLR1	NM_003263	7096
TLR3	NM_003265	7098
TLR4	NM_003266	7099
TLR6	NM_006068	10333
TNFRSF10A	NM_003844	8797
TNK1	NM_003985	8711
TOPK	NM_018492	55872
TPK1	NM_022445	27010
TRAD	NM_007064	11139
TRB2	NM_021643	28951
TRIM	NM_016388	50852
TRIO	NM_007118	7204

TRPM6	NM_017662	140803
TRPM7	NM_017672	54822
TSKS	NM_021733	60385
TTBK1	XM_166453	84630
TTBK2	NM_173500	146057
TTK	NM_003318	7272
TTN	NM_133378	7273
TXK	NM_003328	7294
TXNDC3	NM_016616	51314
TYK2	NM_003331	7297
TYRO3	NM_006293	7301
UCK1	NM_031432	83549
UGP2	NM_006759	7360
ULK1	NM_003565	8408
ULK2	NM_014683	9706

UMP-CMPK	NM_016308	51727
UMPK	NM_012474	7371
URKL1	NM_017859	54963
VRK1	NM_003384	7443
VRK2	NM_006296	7444
VRK3	NM_016440	51231
WEE1	NM_003390	7465
WIF1	NM_007191	11197
XYLB	NM_005108	9942
YES1	NM_005433	7525
YWHAH	NM_003405	7533
YWHAQ	NM_006826	10971
ZAK	NM_016653	51776
ZAP70	NM_001079	7535

Chapter 4

Functional Proteomic Profiling Subdivides Estrogen Receptor-Negative Breast Cancer Into Four Distinct Prognostic Groups

4.1 Abstract

Breast cancer is marked by a deregulation of intracellular and extracellular signaling networks which lead to aberrant cell survival and mitogenesis. Though the critical mitogenic pathways are well described in estrogen receptor alpha (ER)-positive breast cancer, the critical pathways regulating growth of estrogen receptor alpha (ER)-negative breast cancer are still largely unknown. Reverse phase protein microarray technology has the potential to identify differentially expressed and activated proteins and phosphoproteins in large numbers of samples simultaneously. This proteomic approach facilitates the identification of proteins that are differentially expressed and activated by phosphorylation in a way that traditional transcriptional profiling does not allow. We used this proteomic profiling approach to investigate protein expression and activation status in a large panel of human breast tumors. Analyzing the results of these experiments, we identified proteins that are overexpressed and activated in ER-negative breast cancer as compared to ER-positive breast cancer. Furthermore, we demonstrate that ER-negative tumors can be subdivided into four distinct subgroups based on their expression of these proteins, and that these different subgroups have distinct prognostic profiles. We also identified protein signatures that are associated with particularly poor prognosis. Finally, we correlated specific proteomic signatures with previously described breast cancer subtypes identified by transcriptional profiling in human breast cancers. These studies identified proteins and pathways that are

activated in specific subsets of ER-negative breast cancers that can now serve as targets of future drug development for effective treatment of ER-negative breast cancer.

5.2 Introduction

Recent efforts to identify the aberrant signaling pathways that lead to breast cancer transformation and cause breast cancer have focused on identifying DNA or RNA changes in breast cancer. Indeed, studies using comparative genomic hybridization (CGH) arrays and transcription profiling have provided valuable insight into the molecular abnormalities that cause the development of breast cancer [1-6]. These high-throughput technologies have allowed for the evaluation of gene expression or DNA copy number changes on a genome-wide basis and are used to subtype cancers, predict prognosis, and select optimal treatment [3, 5-10]. They have also greatly accelerated cancer drug development [11]. However, by measuring only DNA or RNA changes, these assays are unable to assess the regulation, modification, and activation of proteins, which ultimately effect change in the cell. The development of reliable proteomic characterization techniques is critical for the identification of proteins and pathways that are expressed, activated, and important in human breast cancer.

Proteomics has the potential to complement and further enlarge the wealth of information generated by genomics in breast cancer. It has long been appreciated that mRNA levels do not necessarily correlate with protein abundance [12-14]. Additional complexity is conferred by protein post-translational modifications, including phosphorylations, acetylations, and glycosylations, or protein cleavages (reviewed in [15]). These modifications are not detectable at the mRNA level but play significant

roles in regulating protein stability, localization, interactions, and functions. Finally proteins represent more accessible and relevant therapeutic targets than nucleic acids.

A variety of techniques exist to measure and quantitate protein levels in the cell. Tissue microarrays, protein microarrays, 2-D gel electrophoresis, and mass spectrometry approaches have all been used to probe the proteome in cells and cancers. Tissue microarrays allow for the molecular information to be obtained in the context of cell morphology and tissue architecture, and have been used to query bladder, prostate, colorectal cancers [16-18]. The tissue microarray technique is, however, limited by the need for pathologist scoring and concerns about how representative a small sample (0.6 mm in diameter) is in a potentially heterogeneous tumor. 2-D gel electrophoresis and mass spectrometry approaches have also been used to evaluate protein expression in tumors [19-21]. SELDI-TOF-MS is now routinely used to interrogate protein expression in cells, serum, and tumors. These techniques are limited by the difficulty separating complex protein mixtures and detecting tumor-specific protein traces within a large amount of nonspecific protein species, particularly in a screening or early diagnosis setting where tumor burden is expected to be minimal. Additionally, the need for large starting sample quantities and the labor intensive nature of the techniques limit their clinical utility.

Proteomic arrays depend on immobilizing various protein probes (or protein lysates in the case of reverse phase proteomic arrays) onto specific surfaces and then measuring interactions with specific proteins in complex samples. Recent advances in

robotics and antibody production have allowed for proteomic arrays to subgroup breast cancers [22-28]. Using a novel quantitative protein detection system termed “reverse phase protein arrays” (RPPAs) that relies on validated high-quality antibodies, expression levels and functional activation states of many signaling pathways can now be defined. This technique has provided a novel way to subclassify leukemias [26] and ovarian cancers [25]. RPPA is also able to quantitate very small amounts of protein expression (femtograms of target in nanograms of starting material), and in particular the activation state of cellular signaling pathways and networks using phospho-specific antibodies. Thus, reverse phase protein arrays may be useful for target discovery in addition to being a means of measuring the global activation status of multiple signaling pathways at one time in individual tumor samples.

Estrogen receptor alpha-positive (hereafter referred to as ER-positive) breast cancers account for 60-70% of breast cancers, but the remaining 30-40% of breast cancers are ER-negative and are poorly responsive to traditional therapies [29]. Selective estrogen receptor modulators (SERMs), such as tamoxifen and raloxifene, and aromatase inhibitors are currently used to treat ER-positive breast cancer and have been shown to reduce ER-positive breast cancer recurrence by approximately 50% [30]. Several other examples of effective targeted therapies, including development of the monoclonal antibodies trastuzumab (targeting the HER2/*neu* receptor) and bevacizumab (targeting vascular epithelial growth factor), have been shown to be effective in treating breast cancer [31, 32]. Other primary treatments include small

molecule tyrosine kinase inhibitors including gefitinib and erlotinib (both of which target the epidermal growth factor receptor), and lapatinib (a dual kinase inhibitor targeting both the epidermal growth factor receptor and the HER2/*neu* receptor)[33-36]. These drugs are now being tested in clinical trial and may in the future be used to treat ER-negative breast cancer. These agents, however, are not effective in treating ER-negative breast cancer which don't overexpress HER2. Current treatment options for these tumors are limited to chemotherapy [37]. Such therapy is generally toxic and is not specifically targeted to ER-negative breast cancer cells. Thus, targeted agents that are effective are critically needed for the treatment of ER-negative breast cancer.

To identify novel targets for the treatment of ER-negative breast cancer, including the aggressive ER-negative, PR-negative, HER2/*neu*-negative ("triple-negative") breast cancers, I analyzed data from reverse phase protein arraying (RPPA) experiments done by Dr. Gordon Mills and Dr. Bryan Hennessy at M.D. Anderson Cancer Center to identify proteins and phosphoproteins that are differentially expressed between ER-positive and ER-negative breast cancer. Using RPPA to interrogate the global activation status of approximately 90 proteins and pathways in 166 human breast tumors, my analysis identified 40 proteins and phosphoproteins that were differentially expressed or activated between ER-positive and ER-negative breast tumors. Next, my analysis demonstrated that ER-negative tumors could be subdivided into four distinct subgroups (ER-low, stathmin high, S6 kinase-activated, and HER2-activated) based on the expression of these 40 differentially expressed proteins. We also identified protein

signatures that were associated with a particularly poor prognosis. We then investigated whether this list of differentially expressed proteins and phosphoproteins was able to distinguish ER-positive and ER-negative breast cancers in a larger validation set of 712 human breast tumors profiled and provided by Dr. Gordon Mills at M.D. Anderson Cancer Center. Analysis demonstrated that this list is again robust at distinguishing ER-positive and ER-negative tumors in an unsupervised manner. We also demonstrated that this proteomic signature separated ER-negative tumors into the 4 subgroups identified in the training set (ER-low, stathmin high, S6 kinase-activated, and HER2-activated), and that these different subgroups had distinct prognostic profiles. These results identified proteins and pathways that are activated in specific subsets of ER-negative breast cancers that now serve as targets of future drug development for effective treatment of ER-negative breast cancer.

4.3 Results

To identify potentially important proteins and pathways that are differentially expressed or activated in ER-negative breast cancers, we designed a study to compare protein and phosphoprotein expression levels in ER-positive and ER-negative human breast tumor samples. A summary of the study design is outlined in **Figure 4.1**. Briefly, we measured the expression of 58 proteins and 31 phosphoproteins in 2 sets of human breast cancers. The first set, hereafter referred to as the "training set", contained 166 human breast tumors and was used to identify those proteins and phosphoproteins that were more highly expressed in ER-negative tumors as compared to ER-positive tumors. We then investigated whether this list of differentially expressed proteins and phosphoproteins could segregate ER-positive from ER-negative tumors in an unsupervised manner in a second set of human tumors. This second set, hereafter referred to as the "validation set", was much larger and contained 712 human breast tumors. Because the validation tumor set has comprehensive clinical follow-up data, we were able to use this sets to evaluate time to recurrence, metastasis-free survival, and overall survival. All the tumors were collected by investigators in Denmark and at M.D. Anderson Cancer Center under the direction of Dr. Gordon Mills, and the data for this work was generously provided to us by Dr. Mills for further analysis. All analyses in this chapter were done by Corey Speers with input from Dr. Susan Hilsenbeck at Baylor College of Medicine, unless otherwise noted.

Figure 4.1- Overview of the study design

Figure 4.1 **Human Breast Tumor Samples**

Analysis of 878 human breast tumor samples using reverse phase protein lysate arrays

↓

Training Set of Human Breast Tumor Samples

166 human breast tumors with gene expression and reverse phase protein lysate arrays performed in the laboratory of Dr. Gordon Mills at M.D. Anderson. Training set identifies 40 proteins and phosphoproteins differentially expressed between ER-positive and ER-negative tumors. (permutation *p*-value <.01). Unsupervised hierarchical clustering using differentially expressed protein values identifies 4 subtypes of ER-negative breast cancer

↓

Kruskal-Wallis analysis

Identification of proteins and phosphoproteins associated with the previously identified subtypes of human breast cancer

↓

Validation Set of Human Breast Tumor Samples

712 human breast tumors with reverse phase protein lysate arrays performed in the laboratory of Dr. Gordon Mills at M.D. Anderson. 40 differentially expressed proteins and phosphoproteins from training set validated in this larger tumor set. Same 4 subtypes of ER-negative breast cancer identified in the training set also identified in this validation set.

↓

Kaplan-Meier analysis in validation set shows that HER2 and S6 kinase groups of ER-negative breast cancer have poor overall survival compared to ER-low and Stathmin high groups of ER-negative breast cancer

↓

Identified potential targets for the treatment of ER-negative breast cancer

4.3.1 Patient Population

A total of 878 patients with invasive breast cancer were included in these studies and these tumors were divided into two sets, a training set and a validation set. A summary of the clinical and demographic features of these tumors are summarized in **table 4.1**. These tumors were collected by investigators in Denmark and at M.D. Anderson Cancer Center under the direction of Dr. Gordon Mills, and the data was generously provided to us for further analysis. Breast biopsies using a core needle were taken before initiation of any treatment and were used in this study. Because the patients did not receive systemic adjuvant or neoadjuvant therapy prior to the biopsy, the results from the proteomic analysis represent basal protein expression in these breast cancers. In the training set, 166 patients were obtained from a cohort of patients in Norway (kindly provided by Dr. Myhre to Dr. Mills). In this set, 126 tumors were ER-positive and 40 were ER-negative by IHC-staining. The majority of the tumors (93%) had at least one lymph node positive for disease and came from women who were all under the age of 70.

The validation set was comprised of 712 tumors assembled from centers at M.D. Anderson Cancer Center (621 tumors) and Baylor College of Medicine (91 tumors). In this set, 449 tumors were ER-positive and 263 were ER-negative. The women whose tumors comprised this set tended to be older than the women in the training set (age mean 60.3 and 54.8, respectively) but again, the majority of these women had lymph

Table 4.1- Characteristics of the 166 patients and tumors used in the "training" set and the 712 patients and tumors in the "validation" set to identify differentially expressed proteins and phosphoproteins in human breast cancer. These tumors were acquired by Dr. Gordon Mills at M.D. Anderson and were processed in his laboratory, with data kindly provided for further analysis. SD refers to standard deviation.

Table 4.1. Clinical characteristics of the patients and tumor samples used in the study.

Characteristic		Training Set N=166 (%)	Independent Validation Set N= 712 (%)
Age	<40	11 (7%)	57 (8%)
	40-49	39 (23%)	88 (12%)
	50-59	55 (33%)	112 (16%)
	60-69	63 (37%)	145 (20%)
	>70	0 (0%)	174 (24%)
	Mean	54.8 (SD 9.14)	60.3 (SD 14.9)
	Range	30-69	23-89
Tumor Stage	T1	49 (30%)	181 (25%)
	T2	97 (58%)	268 (38%)
	T3	20 (12%)	103 (15%)
	Unknown	0 (0%)	160 (22%)
Nodal Status	0 nodes positive	11 (7%)	270 (38%)
	1-3 nodes positive	77 (46%)	198 (28%)
	>3 nodes positive	78 (47%)	78 (11%)
	Unknown	0 (0%)	166 (23%)
Tumor Type	Ductal	132 (80%)	411 (58%)
	Non-Ductal	34 (20%)	164 (22%)
	Unknown	0 (0%)	137 (20%)
Recurrence Status	No local recurrence	124 (75%)	363 (51%)
	Local recurrence	42 (25%)	216 (30%)
	Unknown	0 (0%)	133 (19%)
Metastasis	No distant metastasis	67 (40%)	359 (50%)
	Distant metastasis	99 (60%)	216 (30%)
	Unknown	0 (0%)	137 (20%)
Survival Months- All	Mean	107.5 (SD 80.65)	69.86 (SD 58.61)
	Range	6-258	1-265.4
Survival Months- ER-positive	Mean	119.3 (SD 77.5)	80.32 (SD 58.19)
	Range	6-258	1-268
Survival Months- ER-negative	Mean	70.25 (SD 80.1)	54.15 (SD 58.05)
	Range	6-250	2.5-236.5
Molecular Profile Subtype	Luminal A	47 (28%)	NA
	Luminal B	28 (17%)	NA
	ErbB2	42 (25%)	NA
	Normal	24 (15%)	NA
	Basal	25 (15%)	NA
ER	Positive	126 (76%)	449 (63%)
	Negative	40 (24%)	263 (37%)
	Unknown	0 (0%)	0 (0%)
HER2/neu	Positive	34 (21%)	21 (3%)
	Negative	102 (61%)	148 (21%)
	Unknown	30 (18%)	543 (76%)

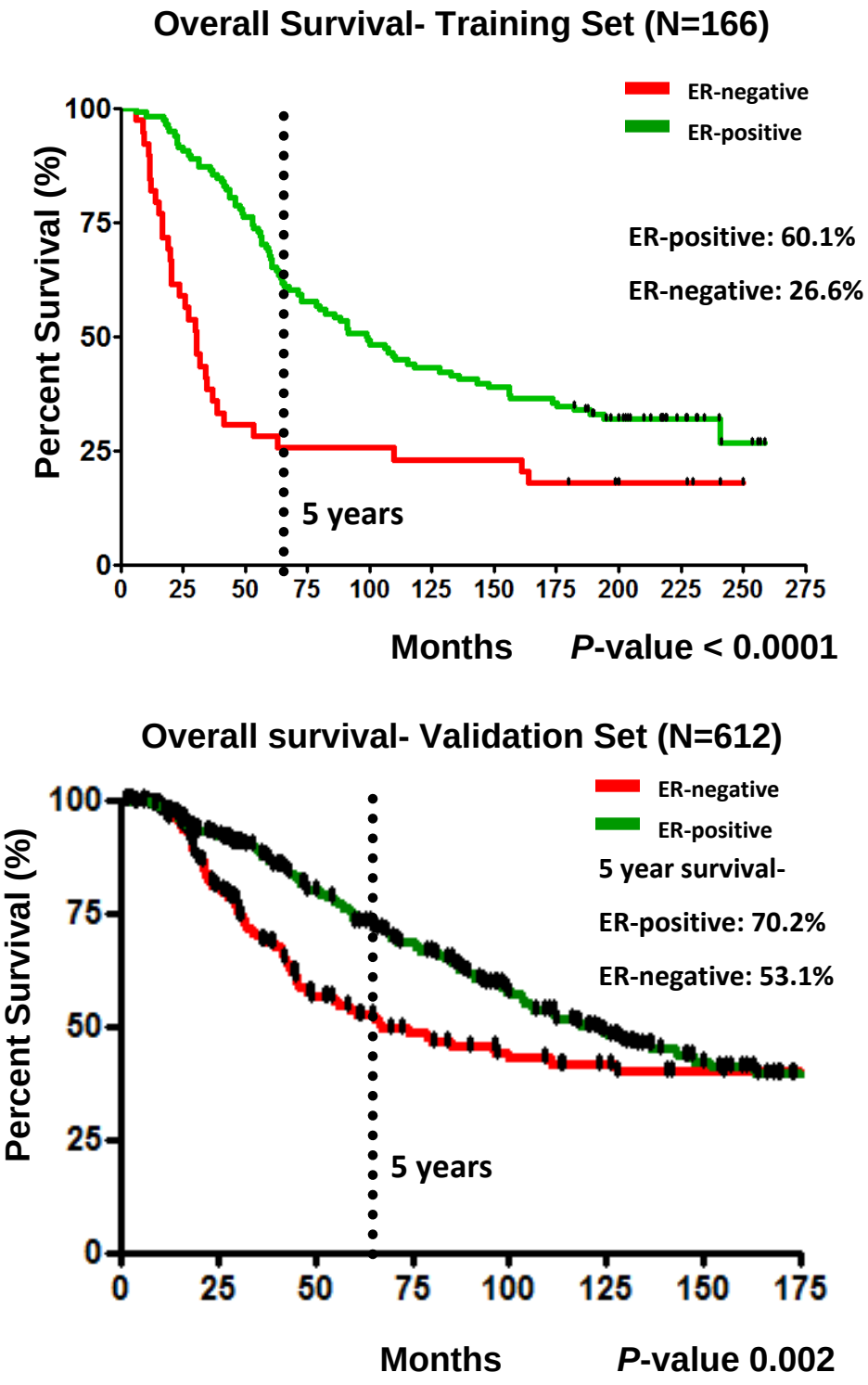
node positive disease (60%). In addition, the tumor samples obtained at M.D. Anderson had clinical follow-up data including date of diagnosis, disease and recurrence free survival, metastasis-free survival, and overall survival). The tumor samples obtained at Baylor College of Medicine did not have clinical follow-up data and thus were censored from any analysis involving clinical follow-up data. The overall survival curves of patients in the two datasets are shown in **Figure 4.2**. The training set had poorer than average overall survival, especially in the patients with ER-negative tumors at 5 years (**Figure 4.2)**. Over 75% of these women died within 5 years, suggesting that these women had particularly aggressive disease. Survival curves from the validation set were more consistent with previously reported curves in which initially patients with ER-negative tumors have a worse overall survival at 5 years, (although survival is similar after 12 years follow-up) (see **Figure 4.2**).

4.3.2 Proteomic Profiling Identified Proteins and Phosphoproteins Overexpressed in Human ER-Negative Breast Tumors

To identify proteins that are differentially expressed in ER-negative breast cancers, we performed reverse phase protein lysate array (RPPA) profiling using 89 antibodies (58 total protein antibodies and 31 phosphoprotein antibodies) to compare human ER-negative and ER-positive breast tumors in the training set. A comprehensive list of the antibodies and sources of antibodies can be found in **Supplementary Table 4.1**. This RPPA technique, described in detail elsewhere [25], has previously been shown

Figure 4.2- Kaplan-Meier analysis shows that ER-negative patients have a worse overall survival in both the training and validation sets at 5 years. (A) Kaplan-Meier analysis of the training set shows that women with ER-negative breast cancer have poorer overall survival at 5 years (26%) compared to ER-positive patients (60%). (B) Kaplan-Meier analysis of the validation set shows that women with ER-negative breast cancer have poorer overall survival at 5 years (53%) compared to ER-positive patients (70%). Outcome data kindly provided by Dr. Gordon Mills at M.D. Anderson.

Figure 4.2


Overall Survival- Training Set (N=166)
ER-negative
ER-positive
ER-positive: 60.1%
ER-negative: 26.6%
Percent Survival (%)
100
75
50
25
0
5 years
0 25 50 75 100 125 150 175 200 225 250 275
Months
P-value < 0.0001
Overall survival- Validation Set (N=612)
ER-negative
ER-positive
5 year survival-
ER-positive: 70.2%
ER-negative: 53.1%
Percent Survival (%)
100
75
50
25
0
5 years
0 25 50 75 100 125 150 175
Months
P-value 0.002

to subcluster leukemias and ovarian tumors into clinically useful subgroups [25, 26]. These experiments were done in the laboratory of Dr. Gordon Mills using his proteomic facilities. Dr. Bryan Hennessy did the proteomic analysis at M.D. Anderson and kindly provided the data for our analysis. In addition to RPPA analysis, we also performed gene expression profiling on this training dataset which allowed us to correlate gene expression changes at the RNA level with protein changes as measured by RPPA. While this study focused on changes identified at the protein level, references to the intrinsic gene set identified by applying the intrinsic gene list [10] are mentioned here and depicted in the **Figure 4.3**.

We used a significance analysis of microarray (SAM) method to identify those proteins and phosphoproteins that were differentially expressed in ER-positive and ER-negative breast tumors. Our analysis revealed a significant difference (permutation *P*-value< 0.01, hereafter referred to as *P*-value) in the expression of 40 proteins between ER-negative and ER-positive tumors with a false discovery rate (FDR) of 1%. To visualize the clustering of the ER-positive and ER-negative tumors, hierarchical clustering analysis was done using only those proteins identified as being differentially expressed between the two groups (**Figure 4.4**). Hierarchical clustering showed that these 40 proteins and phosphoproteins were able to segregate ER-positive and ER-negative tumors. This clustering based on protein expression shows four main clusters of tumors. Two that are ER-positive and two that are ER-negative. One of the ER-positive tumor clusters is marked largely by high expression of ER-alpha (ER), progesterone receptor (PR),

Figure 4.3- Application of the intrinsic gene set to the training set of human tumors identifies the 5 subtypes of human breast cancers (Luminal A, Luminal B, Normal, HER2, and Basal). This classification was used for the Kruskal-Wallis test to identify proteins associated with the breast tumors subtypes.

Figure 4.3

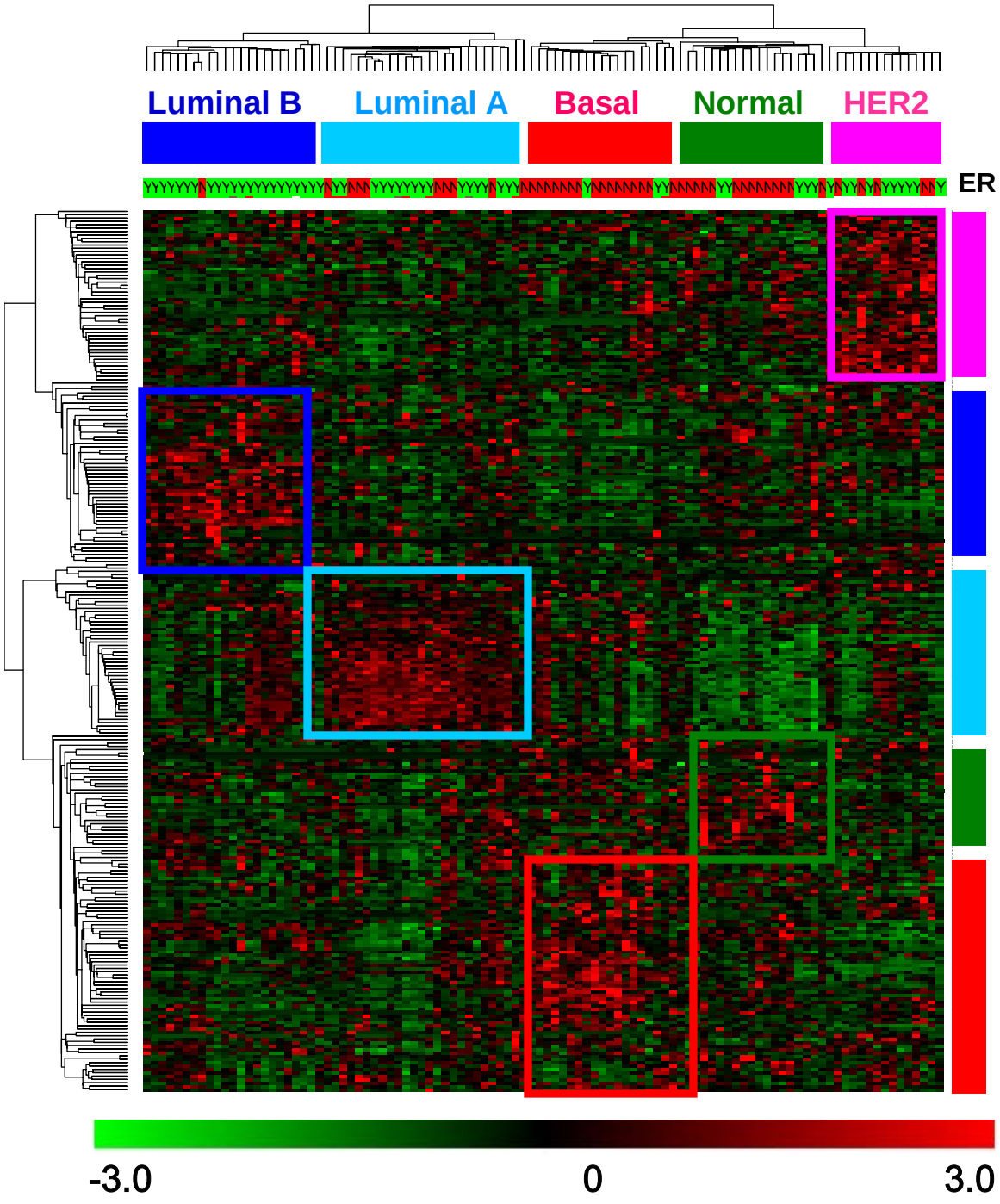


Figure 4.4- Supervised hierarchical clustering of the training set of tumors identifies proteins that are differentially expressed between ER-positive and ER-negative breast tumors. **Supervised hierarchical clustering analysis using the 40 differentially expressed proteins and phosphoproteins accurately clusters the ER-positive from ER-negative human breast tumors. These proteins include those previously known to be associated with ER-status as well as proteins which are associated with pathway activation.**

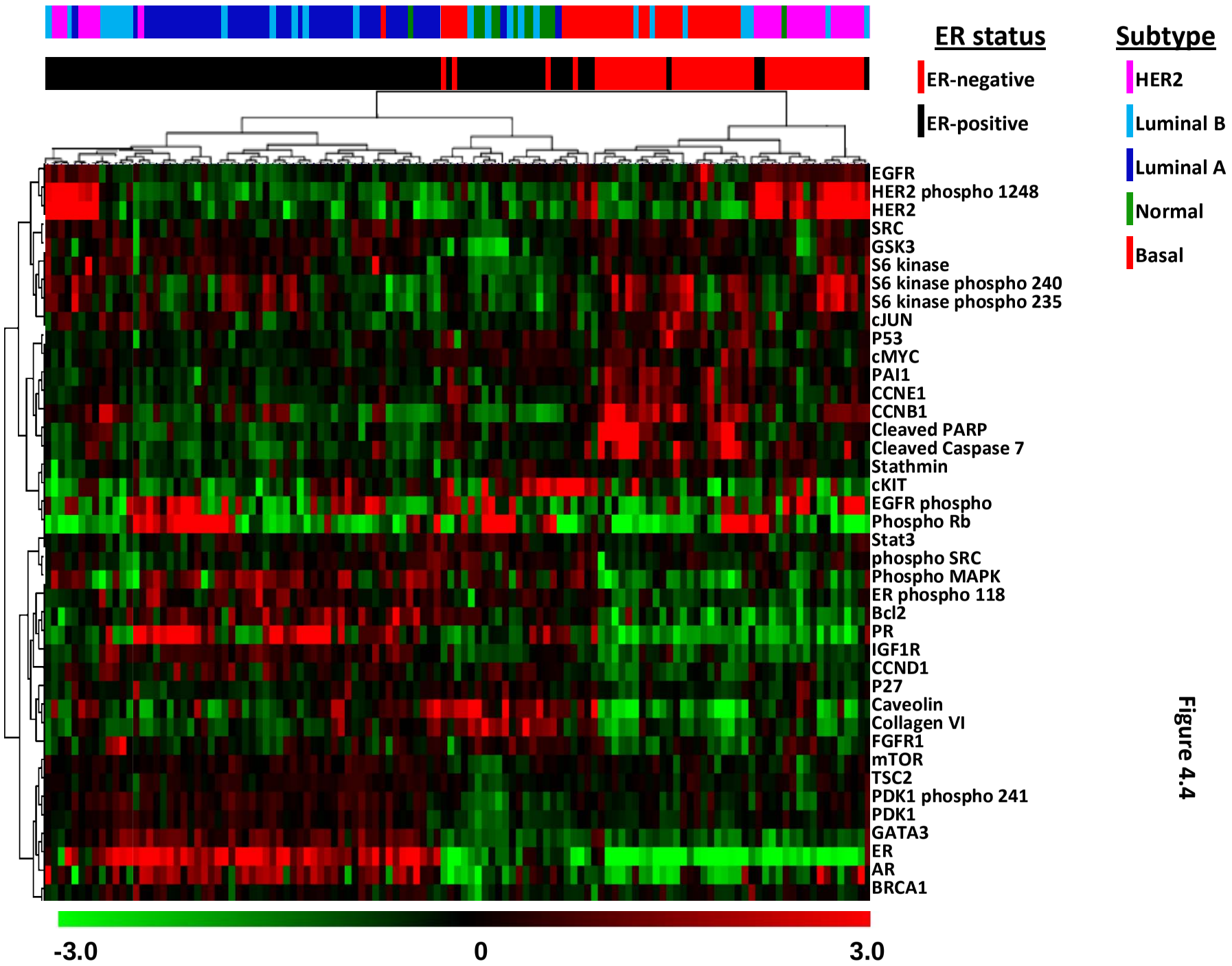

ER status
ER-negative
ER-positive
Subtype
HER2
Luminal B
Luminal A
Normal
Basal
EGFR
HER2 phospho 1248
HER2
SRC
GSK3
S6 kinase
S6 kinase phospho 240
S6 kinase phospho 235
cJUN
P53
cMYC
PAI1
CCNE1
CCNB1
Cleaved PARP
Cleaved Caspase 7
Stathmin
cKIT
EGFR phospho
Phospho Rb
Stat3
phospho SRC
Phospho MAPK
ER phospho 118
Bcl2
PR
IGF1R
CCND1
P27
Caveolin
Collagen VI
FGFR1
mTOR
TSC2
PDK1 phospho 241
PDK1
GATA3
ER
AR
BRCA1
-3.0
0
3.0

Figure 4.4

androgen receptor (AR), GATA binding protein 3 (GATA3), B cell lymphoma 2 (Bcl2), insulin growth factor receptor 1 (IGFR1), phospho MAPK, and breast cancer 1, early onset (BRCA1), amongst others. The samples in this cluster tended to be mostly luminal A tumors, a subset of breast cancers first identified by Sorlie *et al*. [10], though this group did include some ER-positive tumors that were also HER2-positive and thus fell into the gene expression HER2 subtype of breast cancer described previously [10]. The other ER-positive cluster defined by protein expression is marked by tumors with high expression of fibroblast growth factor receptor 1 (FGFR1), caveolin, collagen VI, and cyclin-dependent kinase inhibitor 1B (p27) and is distinct from the first ER-positive cluster of tumors as it has lower relative expression of ER and GATA3 proteins. These tumors were a mix of luminal A, luminal B, basal, and normal-like tumors as determined by application of the intrinsic gene list. Though these tumors are classified as ER-positive by IHC analysis, they represent tumors that express lower levels of estrogen receptor and higher levels of phosphor Rb and phospho EGFR.

The largest ER-negative cluster is marked by complex and heterogenous expression of several proteins including stathmin, c-Src, ribosomal S6 kinase (S6), cyclin E1 (CCNE1), cyclin B1 (CCNB1), plasminogen activator inhibitor 1 (PAI1), cleaved Poly (ADP-ribose) polymerase (PARP), and cleaved caspase 7, and also includes some tumors that express epidermal growth factor receptor 1 (EGFR) and HER2/*neu*. As expected, the majority of these tumors were classified as either basal or HER2 tumors as determined by gene expression profiling (as depicted in **Figure 4.3**), though there was

some luminal B and normal-like tumors as well. The second cluster included tumors that were high in the expression of epidermal growth factor receptor 1 (EGFR), HER2/neu, but also had high expression of S6 and phospho-S6 kinase. These tumors were predominantly classified as HER2 as determined by gene expression profiling. Proteins expected to be overrepresented in ER-negative tumors like EGFR, HER2, and phosphorylated HER2 were identified as being overexpressed in ER-negative tumors. This analysis also identified the S6 signaling pathway, as well as cyclin E and cyclin B, as potentially important targets in ER-negative breast cancer. In addition to levels of total c-Src protein being elevated in ER-negative tumors, phosphorylated c-Src is also elevated, which may identify c-Src as potentially important target in ER-negative breast cancer. A complete list of the proteins differentially expressed between ER-positive and ER-negative breast cancer is shown in **Table 4.2**.

4.3.3 Subtyping of Breast Tumors Using Gene Expression Values and the Intrinsic Gene List Identifies Proteins Associated with Luminal A, Luminal B, Normal-like, Basal, and ErbB2 Breast Tumors.

Recent studies using gene expression profiling have established a widely applied molecular classification of breast cancers [3, 10, 38]. These studies have repeatedly identified 5 subtypes of breast cancer, two luminal-like subtypes (comprised mostly of ER-positive breast tumors and referred to as luminal A and luminal B groups), an erbB2 subtype (comprised mostly of breast cancers, both ER-positive and ER-negative, that

Table 4.2- List of 40 proteins and phosphoproteins that are differentially expressed between ER-positive and ER-negative samples in the "training" set.

Table 4.2. Proteins identified as differentially expressed between ER-positive and ER-negative tumors

Proteins Elevated in ER-negative Tumors	Gene or Protein	Fold Change	Gene Ontology Biological Process Function
Cleaved Poly (ADP-ribose) polymerase	Cleaved PARP	14.4	Apoptosis regulation
Cleaved apoptosis-related cysteine peptidase	Cleaved caspase 7	14.3	Apoptosis regulation
Cyclin B1	CCNB1	4.9	G2/M transition of mitotic cell cycle
HER2/neu	HER2	3.9	transmembrane receptor protein tyrosine kinase signaling pathway
HER2 phospho 1248	Phospho HER2	3.7	transmembrane receptor protein tyrosine kinase signaling pathway
Epidermal growth factor receptor	EGFR	3.7	transmembrane receptor protein tyrosine kinase signaling pathway
Ribosomal S6 kinase phospho (240)	Phospho S6 (240)	3.5	signal transduction, protein kinase activity
Ribosomal S6 kinase phosphor (235)	Phospho S6 (235)	3.4	signal transduction, protein kinase activity
Plasminogen activator inhibitor 1	PAI1	3.2	fibrinolysis, regulation of angiogenesis
Cyclin E1	CCNE1	2.7	G1/S transition of mitotic cell cycle
Ribosomal S6 kinase	S6 Kinase	2.7	signal transduction, protein kinase activity
tumor protein 53	p53	2.7	base-excision repair, protein import into nucleus
v-src sarcoma viral oncogene homolog	SRC	2.5	signal transduction, protein kinase activity
epidermal growth factor receptor phospho (1045)	Phospho EGFR	2.5	transmembrane receptor protein tyrosine kinase signaling pathway
v-src sarcoma viral oncogene homolog phospho (416)	Phospho SRC	2.4	signal transduction, protein kinase activity
stathmin 1/oncoprotein 18	STMN1	2.4	mitotic spindle organization and biogenesis, PI3 kinase activation
cJun	JUN	2.2	transcription factor
v-kit H-Z 4 feline sarcoma viral oncogene homolog	KIT	2.1	protein kinase cascade, signal transduction
glycogen synthase kinase 3	GSK3	2.1	protein kinase cascade, signal transduction, glycogen metabolism

Table 4.2. (continued) Proteins identified as differentially expressed between ER-positive and ER-negative tumors

Proteins Decreased in ER-negative Tumors	Gene or Protein	Fold Change	Gene Ontology Biological Process Function
Estrogen receptor alpha	ER	55	Estrogen mediated signaling, transcription factor
Progesterone receptor	PGR	55	epithelial cell maturation, transcription, signal transduction
Insulin-like growth factor 1 receptor	IGF1R	52	transmembrane receptor kinase signaling, anti-apoptosis
GATA binding protein 3	GATA3	4.9	cell fate determination, transcription
B-cell CLL/lymphoma 2	Bcl2	4.9	humoral immune response, anti-apoptosis
Caveolin 1	CAV1	4.8	inactivation of MAPK activity, vasculogenesis
MAPK phospho	Phospho MAPK	4.5	signal transduction, protein kinase activity, mitogenesis
Collagen VI	COL6A1	4.2	Cell adhesion
Estrogen receptor phospho (118)	ER phospho 118	4.1	Estrogen mediated signaling, transcription factor
Fibroblast growth factor receptor 1	FGFR1	2.9	MAPKKK cascade, cell growth
Cyclin D1	CCND1	2.7	G1/S transition of mitotic cell cycle
Androgen receptor	AR	2.7	Trabscription, signal transduction
breast cancer 1, early onset	BRCA1	2.7	cell cycle checkpoint, DNA repair, DNA replication
signal transducer and activator of transcription 3	STAT3	2.5	regulation of transcription, DNA-dependent, cell motility
Retinoblastoma protein phospho	Phospho Rb	2.5	G1/S transition of mitotic cell cycle, negative reg. of transcription
v-myc myelocytomatosis viral oncogene homolog	MYC	2.1	DNA fragmentation during apoptosis, regulation of transcription
cyclin-dependent kinase inhibitor 1B (p27, Kip1)	P27KIP1	2.1	Cell cycle arrest, cyclin-dependent protein kinase inhibitor activity
pyruvate dehydrogenase kinase, isozyme 1	PDK1	2.0	small GTPase mediated signal transduction, glucose metabolism
pyruvate dehydr. kinase, isozyme 1, phospho (241)	Phospho PDK1	2.0	small GTPase mediated signal transduction, glucose metabolism
mammalian target of rapamycin (mTOR)	mTOR	2.0	signal transduction, cell growth, response to nutrients
tuberous sclerosis 2	TSC2	2.0	negative regulation of kinase activity, protein import into nucleus

overexpress HER2), a basal-like subtype (which tend to be the most aggressive with the poorest prognosis and lack appreciable expression of ER, PR, and HER2; the so called "triple-negative" tumors), and a normal-like subtype (comprised of tumors whose gene expression profile is similar to normal breast epithelial tissue). These subtypes are characterized as having distinct transcriptional profiles, but more importantly, having distinct patient outcomes.

Because we had both gene expression and protein expression data in the training set, we wanted to see if we were able to identify those proteins and phosphoproteins that were associated with each of these 5 subtypes of breast cancer. By applying the intrinsic gene list described previously [3, 10, 38] we assigned the tumors in the training set into one of the 5 subtypes of breast cancer (see **Figure 4.3**). After assigning tumors into one of the five subtypes of human breast cancer, we identified the proteins and phosphoproteins whose expression was significantly correlated with one of these subtypes (using Kruskal-Wallis one-way analysis of variance by ranks). Using this technique, we identified 9 proteins and phosphoproteins associated with luminal A tumors, 6 proteins with luminal B, 12 proteins with normal-like, 3 proteins with erbB2 tumors, and 10 proteins with basal-like tumors. A list of the proteins and their association with the different groups listed in **Table 4.3**. Hierarchical clustering of the significantly associated proteins and phosphoproteins identified using this Kruskal-Wallis test is depicted in **Figure 4.5**.

Table 4.3- List of proteins associated with the previously described subtypes of human breast cancer.

Table 4.3. – List of proteins associated with different subtypes of human breast cancer			
Proteins associated with subtypes of breast cancer	Kruskal-Wallis Statistic (H)	*P*-value	Adjusted *P*-value (Benjamini-Hochberg corrections)
Luminal A tumors			
ER	76.6784	4.107825E-15	3.5327297E-13
GATA3 BD	72.19122	3.5860204E-14	1.5419888E-12
AR	47.499954	4.492669E-9	6.439492E-8
bcl2	43.662155	2.7123452E-8	3.3323096E-7
PR	41.677444	6.8454035E-8	7.3588086E-7
ERp118	41.390305	7.8245314E-8	7.4767746E-7
IGF1R	37.13757	5.620918E-7	4.028325E-6
MAPKp	27.860415	3.875839E-5	1.5151006E-4
LKB1	25.29582	1.221518E-4	4.3771064E-4
Luminal B tumors			
PDK1	31.322906	8.087718E-6	3.819905E-5
PDK1p241	28.13063	3.431982E-5	1.4054783E-4
p70S6 Kinase	25.049938	1.3627403E-4	4.6878267E-4
stat3	18.111946	0.0028093283	0.0075500696
AcCoAp	16.986477	0.0045255153	0.010810953
Akt	15.819983	0.007377275	0.01586114
Normal-like tumors			
ckit	31.358723	7.95696E-6	3.819905E-5
Collagen VI	31.229378	8.439325E-6	3.819905E-5
CCND1	28.211899	3.3086344E-5	1.4054783E-4
SGKp	26.986217	5.739606E-5	2.1461137E-4
SGK	20.84895	8.6504855E-4	0.0026569348
stat3p727	18.124868	0.002793913	0.0075500696
JNK	17.286942	0.003986559	0.010389214
VEGFR2	17.061573	0.00438449	0.010810953
COX2	17.008156	0.004484355	0.010810953
JNKp	16.704386	0.0050959396	0.0116916755
FGFR1	16.65237	0.0052085286	0.0116916755
p27	16.609997	0.005302039	0.0116916755
erbB2 tumors			
HER2	47.727	4.0381334E-9	6.439492E-8
HER2p1248	47.517212	4.456397E-9	6.439492E-8
EGFR	35.24769	1.3427375E-6	7.6983615E-6
Basal-like tumors			
CCNB1	50.816437	9.43092E-10	2.7035306E-8
cleaved PARP	39.601665	1.7966485E-7	1.4769213E-6
cleaved caspase 7	39.493526	1.8890854E-7	1.4769213E-6
PAI1	36.14887	8.868587E-7	5.866911E-6
CCNE1	35.713253	1.0838664E-6	6.6580365E-6
S6p240_4	31.993458	5.959009E-6	3.202967E-5
S6p235-236	22.233332	4.7272968E-4	0.0015636443
S6	20.895502	8.477206E-4	0.0026569348
cMYC	19.83501	0.0013420202	0.003979784
p53	15.217132	0.009473762	0.019398656

Figure 4.5- Kruskal-Wallis analysis identifies proteins associated with the breast tumors subtypes identified by gene expression profiling. After accurately classifying tumors into their intrinsic subtype, Kruskal-Wallis analysis identifies proteins that are associated with the intrinsic subsets of human breast tumors. Analysis done by Corey Speers with input from Dr. Susan Hilsenbeck.

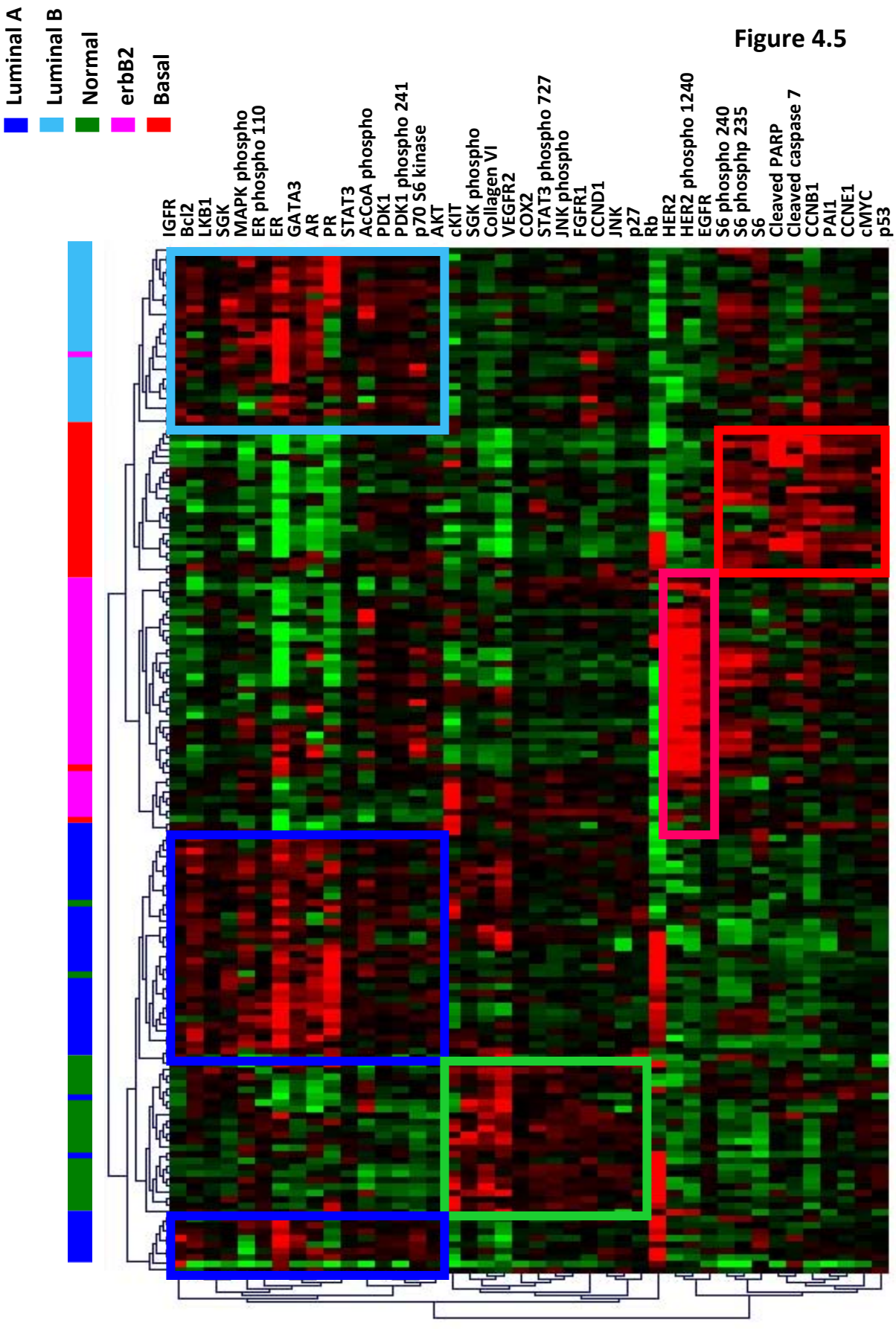

Luminal A
Luminal B
Normal
erbB2
Basal
IGFR
Bcl2
LKB1
SGK
MAPK phospho
ER phospho 110
ER
GATA3
AR
PR
STAT3
AcCoA phospho
PDK1
PDK1 phospho 241
p70 S6 kinase
AKT
cKIT
SGK phospho
Collagen VI
VEGFR2
COX2
STAT3 phospho 727
JNK phospho
FGFR1
CCND1
JNK
p27
Rb
HER2
HER2 phospho 1240
EGFR
S6 phospho 240
S6 phosphp 235
S6
Cleaved PARP
Cleaved caspase 7
CCNB1
PAI1
CCNE1
cMYC
p53

Figure 4.5

4.3.4 Unsupervised Clustering Analysis Reveals Four Distinct Subtypes of ER-Negative Breast Cancer

We next determined whether this list of 40 proteins differentially expressed between ER-positive and ER-negative breast cancers could subcluster the ER-negative tumors in an unbiased manner. We performed unsupervised hierarchical clustering analysis using the 40 proteins and phosphoproteins previously identified as being differentially expressed and found that these tumors clustered broadly into 4 distinct subtypes of ER-negative breast cancer (**Figure 4.6**). Figure of merit analysis showed that these four groups were stable against reclustering. These groups were marked by the coordinated elevation of pathways that fall broadly into 4 different categories.

The first group identified in the training set was small but distinct and had relatively high expression (when compared to the other ER-negative tumors) of proteins that are associated with estrogen receptor positivity, and low expression of the other proteins and phosphoproteins associated with ER-negativity (EGFR, HER2, S6, Src, PAI1, CCNE1, CCNB1). This group, hereafter called the "ER-low" group, is defined by tumors with higher expression of ER alpha, AR, PR, GATA3, Bcl2, and IGF1R than other ER-negative tumors. These tumors fell into the luminal B, HER2, or normal group based on RNA gene expression profiling. The histology and staining of the tumors that fell into this group was reviewed to ensure that they were not misclassified and they were confirmed as negative for ER expression by immunohistochemistry. Additionally, when compared to the other ER-positive tumors in the training set, these tumors expressed

Figure 4.6- Unsupervised hierarchical clustering of the training set tumors identifies 4 distinct subgroups of ER-negative breast cancer. **This unsupervised clustering of ER-negative tumors identifies 4 subgroups of ER-negative tumors and includes a ER-low subgroup, a stathmin subgroup, a HER2 subgroup, and a S6 kinase group. The corresponding breast cancer subtype (luminal A, luminal B, normal-like, HER2, or basal) is noted above the figure. These groupings were derived by applying the intrinsic gene list to gene expression data also collected from these tumors.**

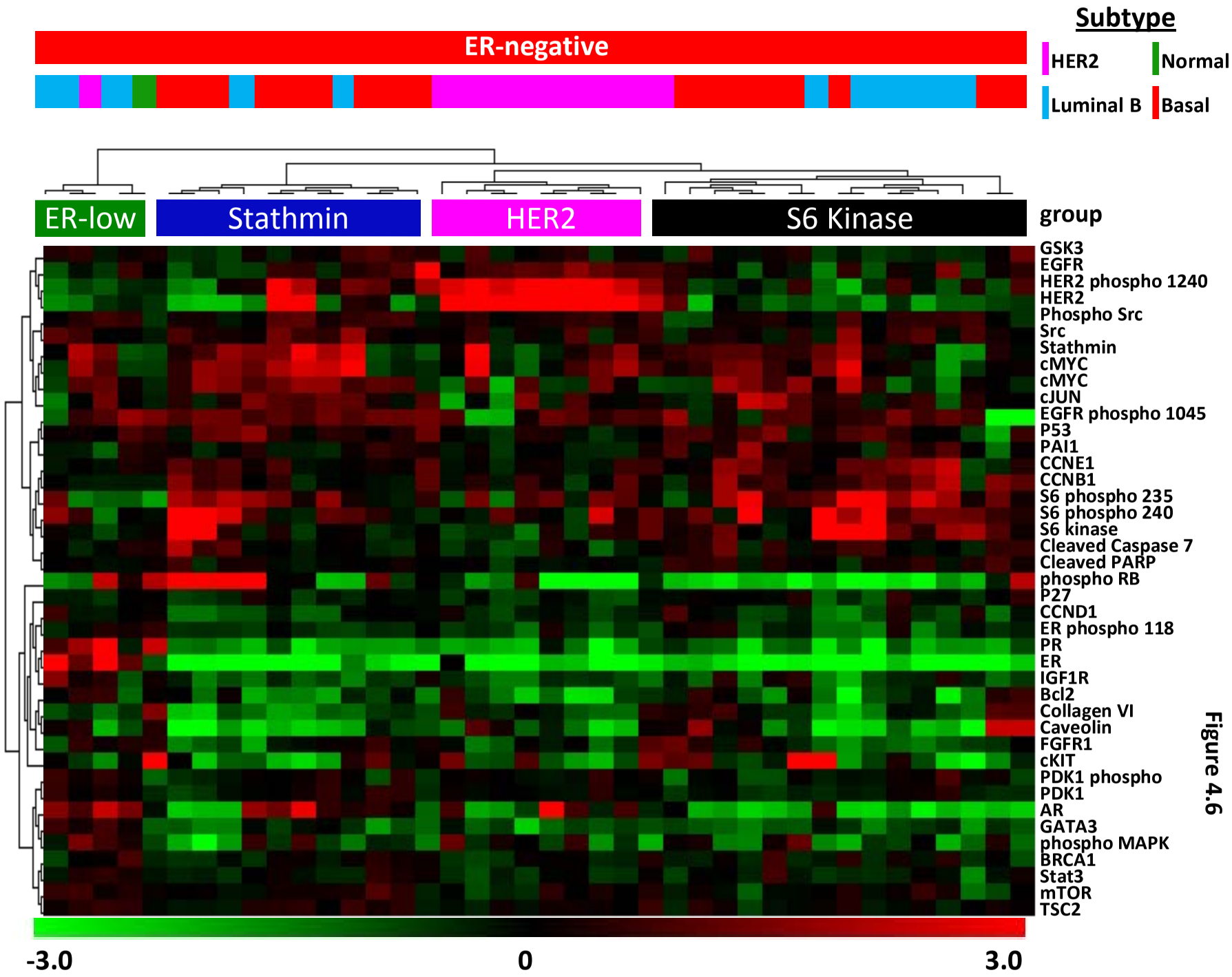

Subtype
ER-negative
HER2
Normal
Luminal B
Basal
ER-low
Stathmin
HER2
S6 Kinase
group
GSK3
EGFR
HER2 phospho 1240
HER2
Phospho Src
Src
Stathmin
cMYC
cMYC
cJUN
EGFR phospho 1045
P53
PAI1
CCNE1
CCNB1
S6 phospho 235
S6 phospho 240
S6 kinase
Cleaved Caspase 7
Cleaved PARP
phospho RB
P27
CCND1
ER phospho 118
PR
ER
IGF1R
Bcl2
Collagen VI
Caveolin
FGFR1
cKIT
PDK1 phospho
PDK1
AR
GATA3
phospho MAPK
BRCA1
Stat3
mTOR
TSC2
-3.0
0
3.0

Figure 4.6

lower levels of ER, PR, and GATA3 protein. Thus, this group represents tumors with low levels of ER expression and activation that is detected by the sensitive RPPA analysis but not detected by less sensitive IHC techniques.

The second group was marked by elevated expression of stathmin (STMN1), Src, phospho-Src, cMYC, and cJUN. This group will be referred to as the "stathmin" group, and correlates to the basal group with some tumors falling into the luminal B group based on RNA gene expression profiling. A third group, hereafter referred to as the "HER2" group, was marked by elevated expression of HER2/*neu*, phospho-HER2/*neu*, EGFR, and phospho-EGFR. This group correlates to the HER2/*neu* subtype of breast tumors identified by intrinsic gene set analysis and is more common in ER-negative breast tumors [4]. The final group was marked by the coordinated high expression of proteins including cyclin B, cyclin E, PAI1, and total and phospho-S6 kinase expression. This group, hereafter referred to as the "S6 kinase" group may identify those tumors that are particularly mitotically active as many of the cyclins (CCNE1, CCNB1), as well as proteins involved in mitogenesis (S6 kinase), are elevated. They are also identified by apparent activation of the S6 kinase signaling pathway. These tumors fall into the basal or luminal B groups based on RNA gene expression profiling. These results indicate that ER-negative tumors may be subdivided into 4 distinct classes based upon their protein expression profile.

4.3.5 Correlation Between RNA and Protein Expression

Because almost all of the tumors in the training set had also been used for previous gene expression profiling experiments, we were able to perform a large scale comparison of RNA and protein expression levels in human breast tumors. Using Spearman's rank correlation test we compared total protein expression with RNA levels for the 166 tumors in the training set, with results reported in **Figure 4.7**. In our hands, for many genes, it appears that RNA and protein levels are not strongly correlated, at least when RNA expression is measured using the Applied Biosystem Human Genome Survey Microarray version 2.0 and protein expression is measured by RPPA. There was a strong correlation between protein and RNA levels (*r* values > .5) in only 9 of 43 genes. There was moderate correlation between protein and RNA levels (*r* values between 0.3-0.5) in 7 of 43 genes, and 27 of 43 genes showed weak or no correlation between protein and RNA levels (*r* values < 0.3) (for all results refer to **Table 4.4**).

4.3.6 Validation of Differentially Expressed Proteins in an Independent Set of Human Breast Tumors

We next used the 40 differentially expressed proteins that we identified in the training set in a validation set of tumors. The validation dataset was comprised of 712 tumors (612 tumors from M.D. Anderson Cancer Center and 91 tumors from Baylor College of Medicine). The tumor samples obtained and processed at M.D. Anderson by Dr. Gordon Mills, and had clinical follow-up data including date of diagnosis, disease and

Figure 4.7- Correlation between RNA and protein expression in the training set. **The RNA and protein expression for many of the genes interrogated is displayed and corresponding correlation coefficients are listed in table 4.4. RNA expression units are log2 transformed values and protein expression units are mean centered relative expression. Analysis done by Corey Speers with input from Dr. Susan Hilsenbeck.**

Figure 4.7

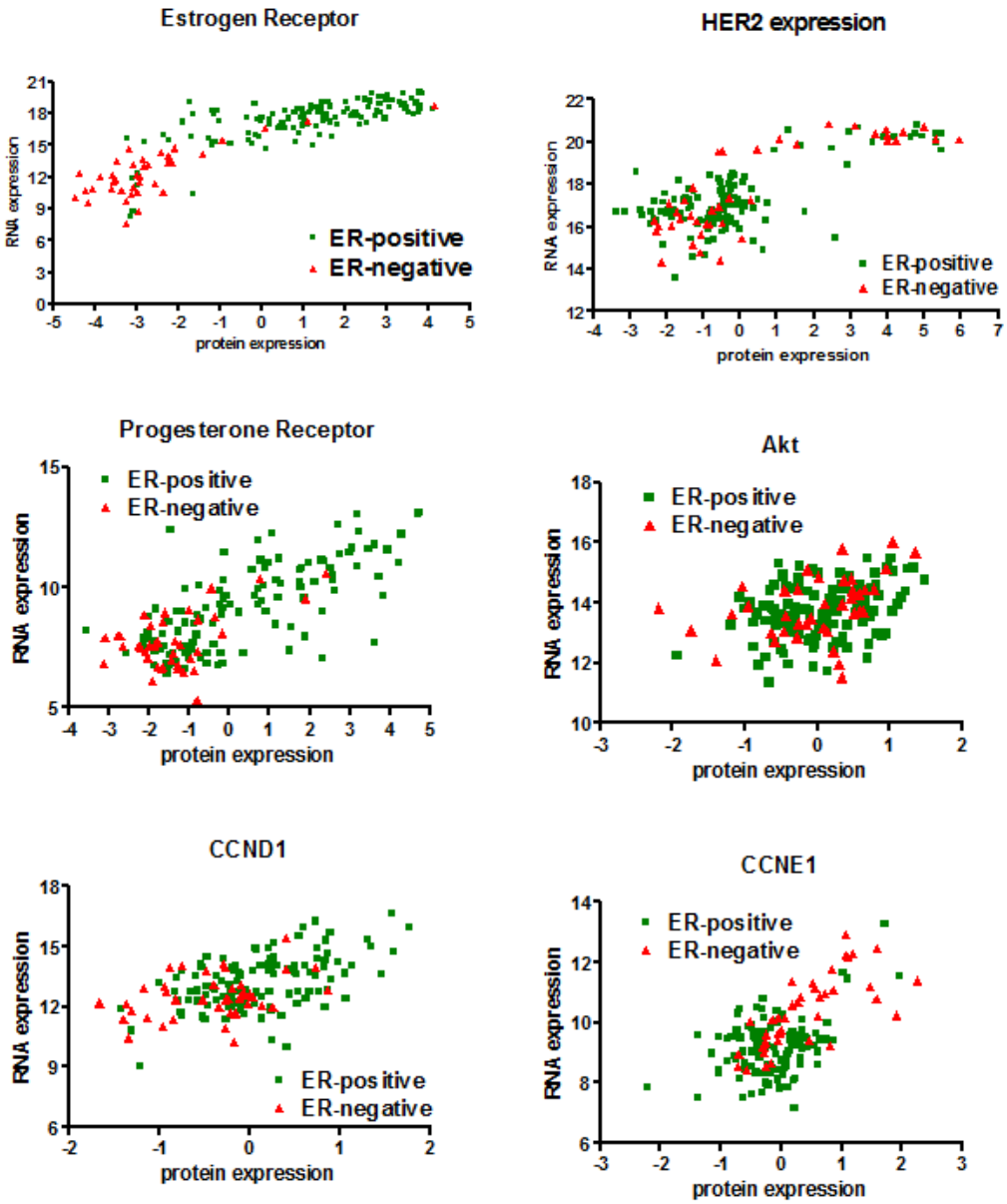

Estrogen Receptor
RNA expression
protein expression
ER-positive
ER-negative
HER2 expression
RNA expression
protein expression
ER-positive
ER-negative
Progesterone Receptor
RNA expression
protein expression
ER-positive
ER-negative
Akt
RNA expression
protein expression
ER-positive
ER-negative
CCND1
RNA expression
protein expression
ER-positive
ER-negative
CCNE1
RNA expression
protein expression
ER-positive
ER-negative

Figure 4.7 (continued)

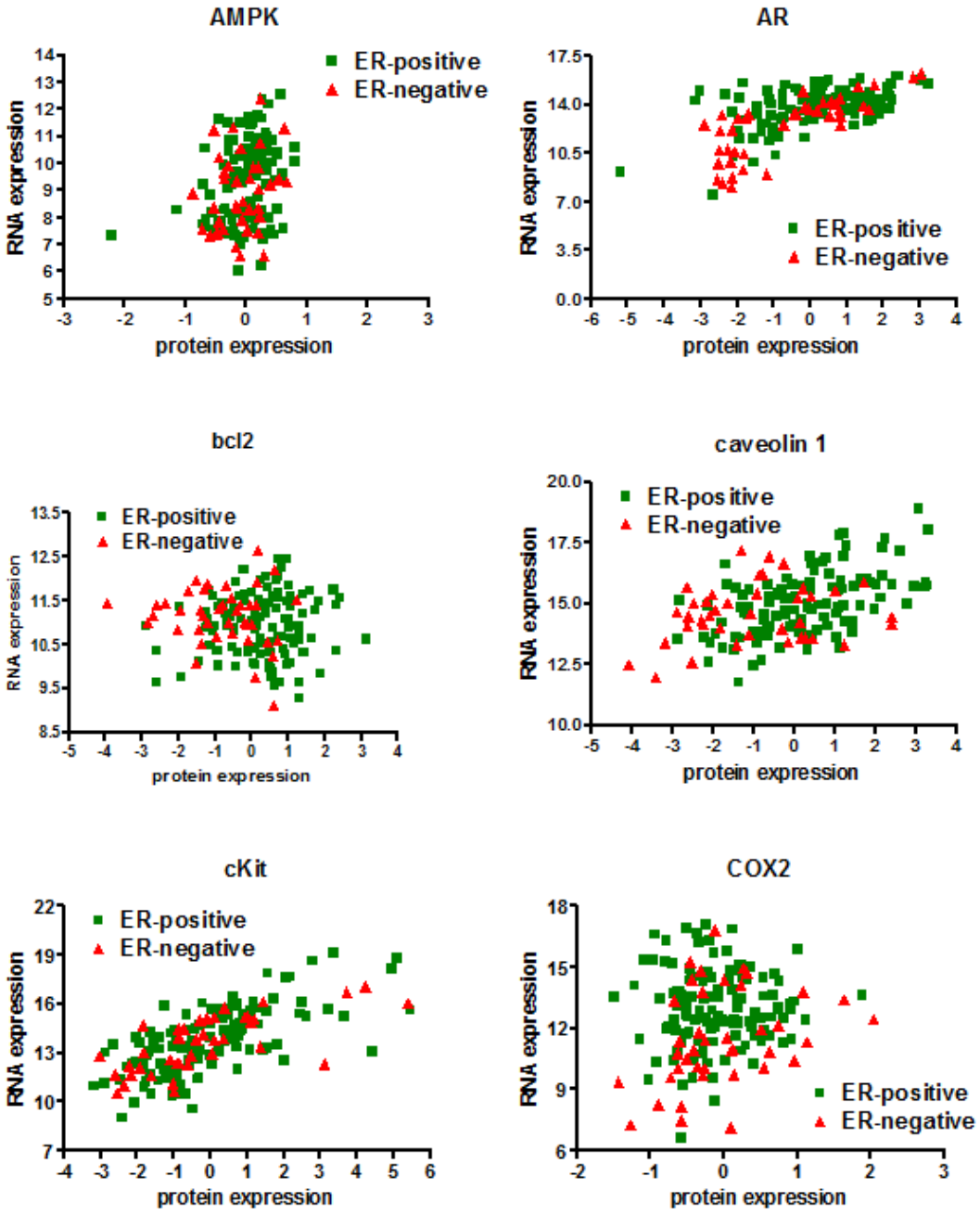


Figure 4.7 (continued)

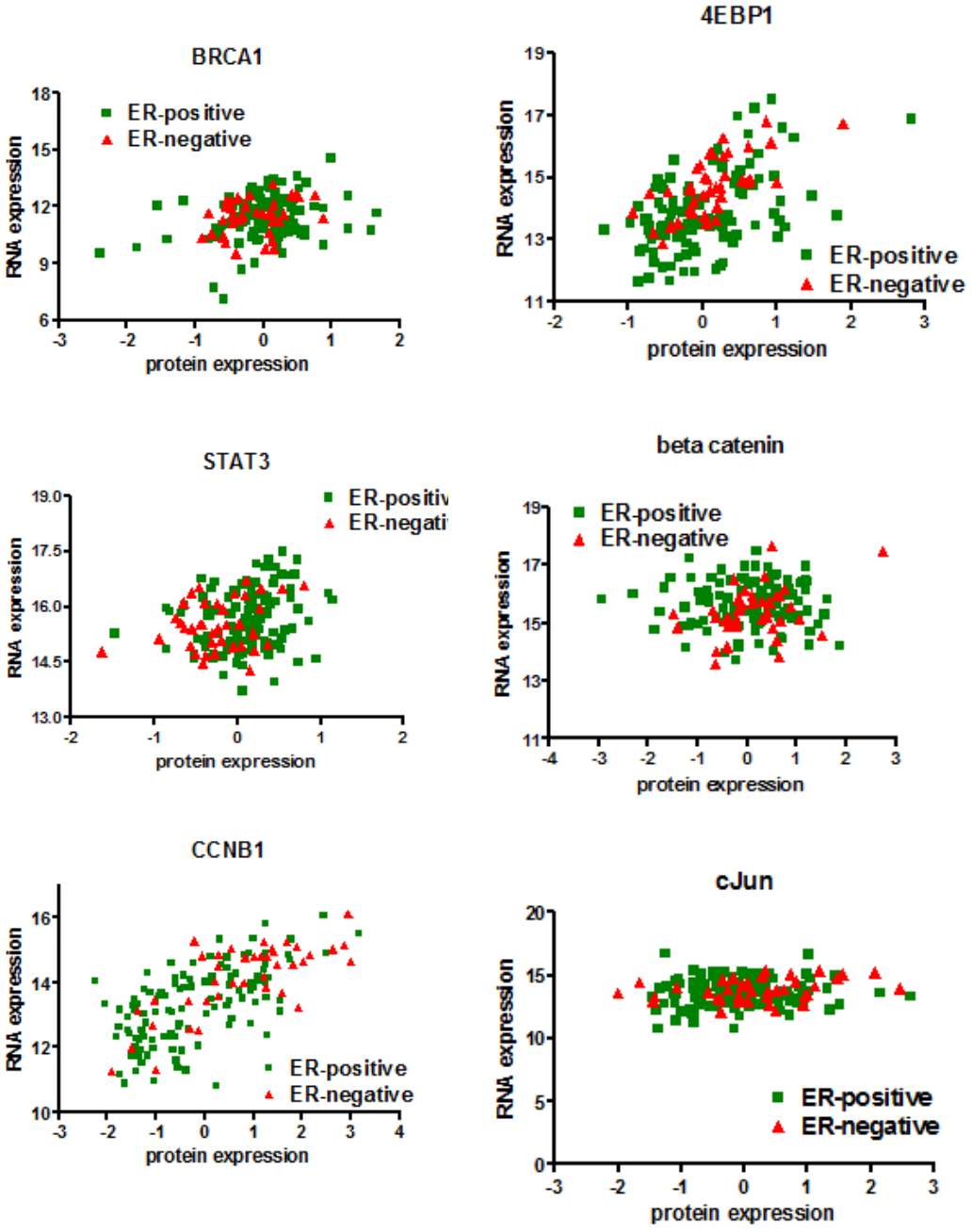


Table 4.4- Correlation r values for the RNA and protein expression in the training set. Analysis done by Corey Speers with input from Dr. Susan Hilsenbeck.

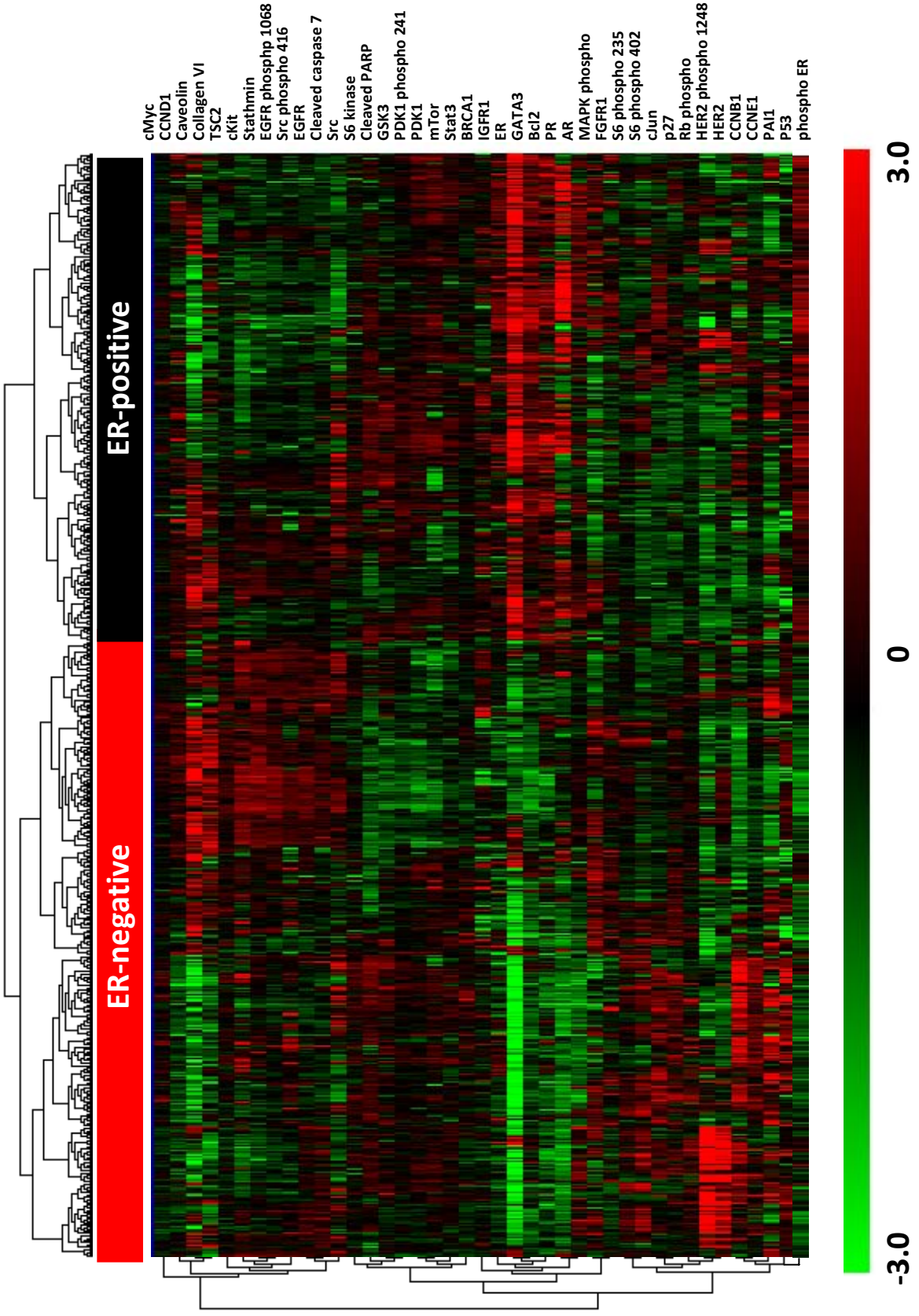
Table 4.4 – Correlation coefficients between RNA and protein expression values in the training set of human breast tumors. Analysis done by Corey Speers with input from Dr. Susan Hilsenbeck.

Protein	Overall *r* value	*P*-value
estrogen receptor	0.85	<0.0001
progesterone receptor	0.69	<0.0001
Cyclin B1	0.69	<0.0001
cKIT	0.69	<0.0001
IGF1R	0.63	<0.0001
GATA3	0.58	<0.0001
HER2	0.57	<0.0001
androgen receptor	0.55	<0.0001
FGFR1	0.54	<0.0001
4EBP-1	0.47	<0.0001
Cyclin D1	0.46	<0.0001
caveolin	0.44	<0.0001
Cyclin E	0.41	<0.0001
EGFR	0.36	<0.0001
cMYC	0.34	<0.0001
Akt	0.31	<0.0001
AMPK1	0.29	0.002
STAT3	0.24	0.002
PTEN	0.21	0.004
Rb	0.18	0.01
BRCA-1	0.17	0.03
MEK1	0.17	0.03
TSC2	0.17	0.018
p53	0.13	0.11
stathmin	0.12	0.06
beta catenin	0.11	0.14
cJUN	0.11	0.15
NOTCH3	0.09	0.25
GSK3	0.09	0.26
XIAP	0.09	0.13
p27	0.08	0.34
COX2	0.07	0.32
PAI1	0.07	0.18
p21	0.06	0.44
LKB1	0.05	0.55
VEGFR2	0.04	0.29
p38	0.03	0.36
p70 S6 Kinase	0.02	0.39
ERK2	-0.014	0.82
Bcl-2	-0.02	0.75
E cadherin	-0.12	0.12
PDK1	-0.12	0.06
SGK	-0.19	0.007

recurrence free survival, metastasis-free survival, and overall survival. The tumor samples obtained at Baylor College of Medicine were processed by Corey Speers and had initial diagnosis data but did not have clinical follow-up data and thus were censored from any analysis involving clinical follow-up data. The clinical characteristics of this dataset were generally similar to those of the training set, though there were differences in the percentage of lymph node-positive patients, HER2 status, and median survival (see **Table 4.1**). Using the 40 differentially expressed proteins and phosphoproteins identified in the training set, we performed unsupervised hierarchical clustering to determine whether these specific proteins and phosphoproteins could accurately segregate ER-positive and ER-negative tumors. Using clustering by Pearson's rank correlation with complete linkage this set of proteins accurately clustered ER-positive and ER-negative tumors in an unsupervised manner (**Figure 4.8**). As with the training set, 4 high order groups were seen: 2 groups of ER-positive tumors and 2 groups of ER-negative tumors. Again the ER-positive tumors showed one group with high expression of ER, PR, AR, GATA3, Bcl2, IGF1R, among others. The other ER-positive group had lower relative expression of ER but had higher expression of Src or caveolin. In the ER-negative tumors, there were again two groups with complex expression patterns. One was a complex group of tumors with high HER2 or CCNB1 or CCNE1 expression. The other complex group has high caveolin and collagen VI expression or high stathmin, EGFR, and Src expression. Notably, all of the proteins and

Figure 4.8- Unsupervised hierarchical clustering in the validation set accurately clusters ER-positive and ER-negative tumor samples based on the previously identified protein signature. **Unsupervised hierarchical clustering analysis of differentially expressed proteins and phosphoproteins identified in the training set show that this signature separates ER-positive and ER-negative tumors in the validation set.**

Figure 4.8


cMyc
CCND1
Caveolin
Collagen VI
TSC2
cKit
Stathmin
EGFR phosphp 1068
Src phospho 416
EGFR
Cleaved caspase 7
Src
S6 kinase
Cleaved PARP
GSK3
PDK1 phospho 241
PDK1
mTor
Stat3
BRCA1
IGFR1
ER
GATA3
Bcl2
PR
AR
MAPK phospho
FGFR1
S6 phospho 235
S6 phospho 402
cJun
p27
Rb phospho
HER2 phospho 1248
HER2
CCNB1
CCNE1
PAI1
P53
phospho ER
ER-positive
ER-negative
3.0
0
-3.0

phosphoproteins identified as being differentially expressed in the training set were significantly differentially expressed in the validation set.

4.3.7 Unsupervised Clustering Analysis Again Reveals Four Distinct Subtypes of ER-Negative Breast Cancer.

Utilizing the same approach we did with the training set, we performed unsupervised hierarchical clustering using only the ER-negative tumors in the validation set. Using Pearson's rank correlation and complete linkage, we again identified four distinct subsets of ER-negative breast cancer (**Figure 4.9**). As with the training set, figure of merit analysis showed that these four groups were stable against reclustering. These groups again included an ER-low cluster, stathmin expressing cluster, an S6 kinase cluster, and a HER2 overexpressing cluster. The clustering of the proteins was also very similar in this dataset, with the ER-associated proteins (ER alpha, PR, AR, GATA3, Bcl2) clustering closely, HER2/*neu* and EGFR clustering closely, S6 and phospho-S6 clustering again with cyclin B and E and PAI1, and stathmin clustering closely with cMYC, Src, and phospho-Src. As with the training set, a small group of "ER-low" tumors were identified, and may represent tumors that have estrogen receptor expression that is too low to be detected by IHC analysis, but are identified as having estrogen receptor expression by the more sensitive RPPA technique.

Figure 4.9- Unsupervised hierarchical clustering of the ER-negative tumors in the validation set. **Using the 40 differentially expressed proteins, unsupervised hierarchical clustering again identifies 4 subsets of ER-negative tumors. These clusters include the ER-low, stathmin, HER2, and S6 kinase groups.**

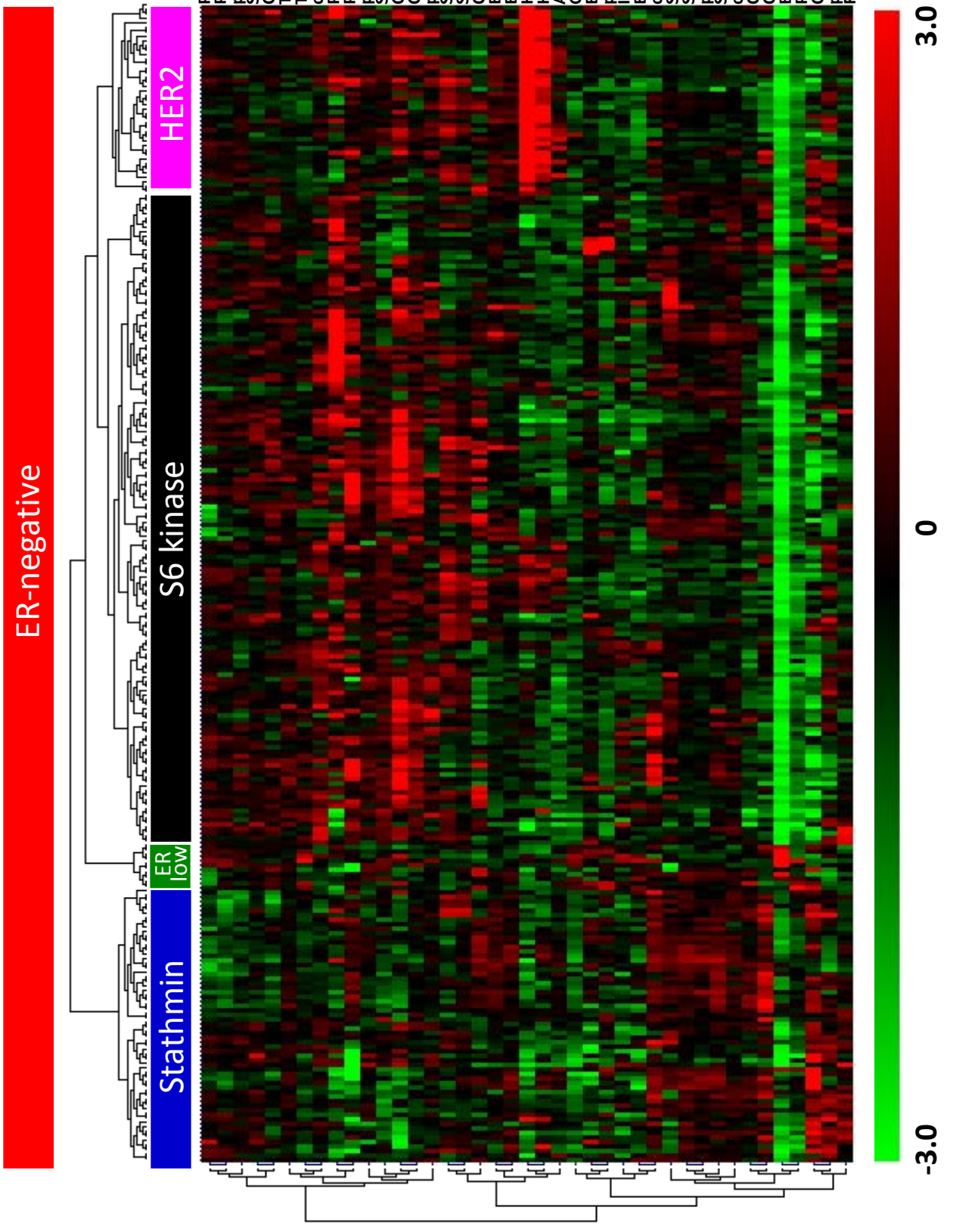

ER-negative
HER2
S6 kinase
ER low
Stathmin
PDK1
PDK1 phospho 241
phospho SGK
Stat 3
Cleaved PARP
TSC2
TSC2
cJun
PAI1
P53 total
p53
S6 kinase
CCNB1
CCNE1
phospho Rb
S6 phospho 240
S6 phospho 235
Cleaved caspase 7
EGFR
EGFR phospho 992
HER2 phospho 1248
HER2
AR
GATA3
BRCA1
phospho SGK
IGFR1
Bcl2
cKit
Src phospho 416
Src
p27
Stathmin
cMYC
CCND1
Collagen VI
ER
PR
Caveolin
phospho MAPK
FGFR1
3.0
0
-3.0

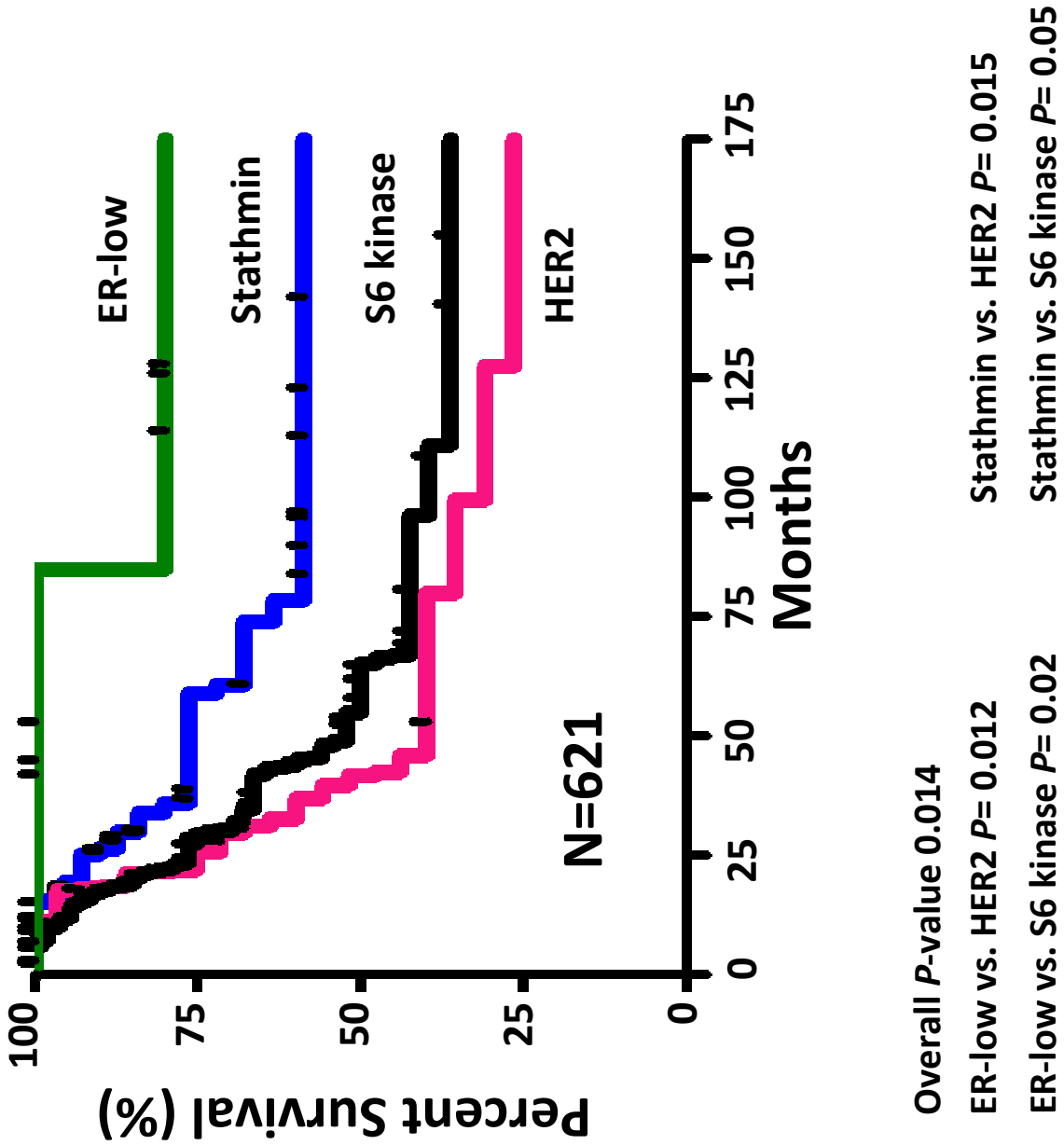
Figure 4.9

4.3.8 S6 Kinase and HER2/*neu* Subtype of ER-Negative Breast Cancer Predicts Poor Overall Survival.

To determine whether any additional insight into these subtypes of ER-negative breast cancer could be ascertained in this larger dataset, we analyzed the survival data from this dataset (only the 621 tumors from M.D. Anderson that had clinical follow-up data). The median follow-up time of these tumors was 96 months Based on the unsupervised clustering of the ER-negative tumors, these tumors were categorized into one of the four subsets: 1) ER-low, 2) stathmin, 3) HER2/*neu*, or 4) S6 kinase tumors. Kaplan-Meier analysis of the overall survival between the different subgroups of ER-negative tumors (ER-low, stathmin, HER2/*neu*, S6 kinase) shows that there were differences in the overall survival curves between the women belonging to these different subtypes. Women with the S6 kinase and HER2/*neu* signature-expressing tumors had significantly worse overall survival than the women in the ER-low or stathmin groups (see **Figure 4.10**). These data validate the 40 proteins and phosphoproteins as being differentially expressed between ER-positive and ER-negative tumors and again demonstrate that protein expression in ER-negative tumors identify women who have distinct clinical outcomes based on their protein expression profile.

Figure 4.10- Kaplan-Meier analysis of overall survival in the ER-negative tumors in the validation set. **Patient's tumors were assigned to one of the 4 subsets of ER-negative breast cancer based on unsupervised hierarchical clustering of the differentially expressed proteins. Analysis shows those patients with activated HER2/*neu* or S6 kinase signaling pathways have reduced survival as compared to those patients with an ER-low or stathmin signaling pathway signature.**

Figure 4.10


Percent Survival (%)
100
75
50
25
0
0
25
50
75
100
125
150
175
Months
N=621
ER-low
Stathmin
S6 kinase
HER2
Overall P-value 0.014
ER-low vs. HER2 P= 0.012
ER-low vs. S6 kinase P= 0.02
Stathmin vs. HER2 P= 0.015
Stathmin vs. S6 kinase P= 0.05

4.4 Discussion

In this report we demonstrate that ER-negative tumors can be subdivided into prognostically distinct subgroups based on their expression and activation of signalling proteins. Additionally, we identify many proteins and pathways that are overexpressed or activated in ER-negative breast cancer. Women whose tumors express low levels of ER or ER-regulated genes (ER-low group) or have high stathmin expression (stathmin group) have relatively good prognoses as compared to women whose tumors have the HER2/*neu* (HER2 group) or S6 kinase pathways (S6 kinase group) activated. These studies underscore the utility of RPPA in assessing the global protein and pathway activation status of human breast tumors and identify proteins and pathways which may be particularly attractive targets for the treatment of ER-negative breast cancer.

This is the first study to use such a large set of human breast tumors for proteomic profiling. This large dataset allows for the identification of smaller but perhaps significant subtypes of breast cancer that may remain undetected in a smaller sample set. The ER-low subgroup of ER-negative breast cancer represents one such subtype. In these datasets, only about 5% of ER-negative tumors fell into this category, but this subgroup may represent a clinically important group of patients who are not currently treated with anti-estrogen therapy (based on negative ER staining by IHC). Based on the low but detectable expression of ER in these patients, they may in fact benefit from such anti-estrogen therapy, and though they fall into a group with

relatively good prognosis, such anti-estrogen therapy may further improve their prognosis.

Having accurate and extensive clinical follow-up data on these tumors allowed for analyses of prognostic biomarkers and the identification of proteins whose expression is correlated with either good or poor overall survival. Additionally, having also transcriptionaly profiled the training set of tumors, we were able to characterize the discordance between the 4 subgroups of ER-negative breast cancer and the groups they cluster into by RNA expression analysis. While RNA gene expression profiling studies show that ER-negative tumors are either normal-like, basal-like, or express HER2, we identify additional, clinically relevant subtypes. This correlation showed that, unlike with RNA expression clustering, basal-like tumors had heterogeneity in their protein expression profiles, as did the normal-like and HER2-overexpressing tumors. We found that not all S6-kinase group tumors were basal, but were a mixture of basal and luminal B tumors. The stathmin high group tumors were basal or luminal B by gene expression profiling, and the ER-low group of ER-negative tumors included multiple subtypes including normal-like, basal, and HER2 based on gene expression profiling. This represents a novel classification based on protein expression, and the differences in clinical outcomes between these groups merits further investigation.

Another advantage of having both RNA and protein expression data, was the ability to identify proteins that are associated with the breast cancer subtypes previously identified using gene expression profiling [4, 10]. We identified several

proteins and phosphoproteins associate with luminal A tumors (ER, GATA3, AR, Bcl-2, PR, ERp118, IGF1R, MAPKp, amongst others). Luminal B tumors were associated with PDK1, PDK1p241, p70S6 Kinase, and Akt expression, amongst others. These proteins, known to interact and phosphorylate each other after mitogen stimulation [39, 40], represent a potential novel targetable pathway in these more aggressive, ER-positive tumors. Future work will need to evaluate whether inhibition of the PDK1/S6 kinase signaling pathway in models of aggressive, ER-positive tumors (luminal B), is an effective treatment strategy.

ER-negative tumors include normal-like, ErbB2-positive, and basal-like tumors. Normal-like tumors were associated with a wide variety of protein and phosphoprotein expression, including cKIT, SGKp, SGK, COX2, and p27 amongst others. ErbB2-positive tumors were correlated with expression of total and phosphor-HER2 and EGFR, known to be overexpressed and dimerization partners in tumors which overexpress the HER2/*neu* gene. Finally, basal-like tumors were associated with expression of CCNB1, cleaved PARP, cleaved caspase 7, PAI1, CCNE1, S6 kinase, and cMYC, as well as others. These proteins are involved in cell cycle regulation as well as apotosis, and may represent two pathways that are particularly active in these aggressive, basal tumors. As patients with basal-like tumors have a poor prognosis and significantly worse outcomes than patients in the other subtypes [29, 38], this list represents potential targets for specific therapy in patients with basal-like tumors.

Proteomic profiling identified 40 proteins that were differentially expressed between ER-positive and ER-negative breast cancer. Many of these proteins have previously been associated with either ER-positivity or ER-negativity. Our results validate previously reported studies, and underscore the ability of RPPA to detect differences at the protein level between different types of breast cancer. As was expected, proteins like estrogen receptor-alpha, progesterone receptor, GATA3, Bcl-2, all previously reported to be elevated in ER-positive breast cancer, were also more highly expressed in our sets of ER-positive breast cancer. The heterogeneity, however, in protein expression amongst the ER-positive tumors wasn't expected. There were groups of ER-positive tumors that had lower relative expression of ER and ER-associated proteins, but had high expression of IGF1R, FGFR1, CCND1, and phospho-MAPK. Indeed, previous studies have identified many of these genes as being under estrogen receptor transcriptional control [41-44]. Estrogen receptor is also known to be a rapid activator of the IGF1R pathway (and IGF activates ER signaling pathways) and the interplay between these signaling networks drive proliferation, metastasis, and defends against apoptosis in breast cancer cell lines [45, 46]. Both CCND1 and FGFR1 can be amplified in ER-positive breast cancer and may be involved in endocrine therapy resistance [47-49]. Interestingly, the identification of activated MAPK in ER-positive breast cancer may again identify an important pathway involved in endocrine therapy resistance. This alternate signaling pathway is currently an area of intense investigation and trials are

currently underway to determine whether inhibition of MAPK signaling in combination with anti-estrogen therapy will overcome such resistance [50-52]

There has been much debate about whether gene expression profiling is an accurate surrogate measurement with relevance to biology as there are many post-transciptional levels of control which may affect cellular physiology. As proteins are responsible for transduction of cellular signals, attempts have been made to determine whether RNA and protein levels correlate. These previous reports have been conflictory, with some reporting strong correlations, while others report much lower correlation between RNA and protein expression [12, 14, 53, 54]. In this study RNA and protein levels did not correlate very well for many proteins measured (27 or 43), at least when RNA expression is measured using the Human Genome Survey Microarray platform (Applied Biosystems) and protein expression is measured using RPPA. Why, then, does RNA expression profiling give biologically relevant and robust profiles that can be validated in multiple datasets? While it is true that RNA and protein levels do not correlate for all genes, many do and it is these genes that may be the drivers of biologically relevant clusters. Furthermore, despite discrepancies in RNA and protein expression, RNA levels may actually reflect the activation of other upstream proteins that regulate important biological functions. Thus, though the RNA and protein levels a particular gene may not correlate well, it may represent a biomarker of a pathway that is truly important in disease. Finally, RNA may itself as a regulator of cellular function (examples include small interfering RNAs and microRNAs) which modulate gene

expression and translation. RNA levels may also be sensed by cellular machinery that can activate pathways regulating mitogenesis, cell cycle entry, and apoptosis while the RNA itself is never translated into protein. Thus, though RNA and protein levels would not correlate in such a circumstance, there would be reason for RNA levels to influence the biology of these tumors.

The long term goal of these studies is to identify novel targets for the treatment of ER-negative breast cancer. While HER2/*neu* is an important target for those ER-negative tumors that express HER2/*neu*, there are no clear targets for the remainder of ER-negative, PR-negative, HER2-negative tumors. ER-negative breast cancer is currently defined in terms of the proteins that it does NOT express. This work alters the paradigm and identifies genes and proteins that ARE expressed in these tumors. It is no longer necessary to describe these tumors as lacking the expression of estrogen receptor, progesterone receptor, or HER2. Instead, these tumors can be described as expressing an activated S6 kinase pathway, Src, or stathmin. Such markers can subgroup these tumors and are putative targets of treatment. The identification of these proteins and pathways that are overexpressed or activated in ER-negative breast cancer represents an important first step. Besides previously known proteins including EGFR, HER2, and cKIT, these studies identify CCNB1, CCNE1, Src, PAI1, cJUN, stathmin, and GSK3 as proteins highly expressed and activated in ER-negative tumors. Future studies should focus on determining which of these proteins and pathways are critical for growth,

invasion, and metastasis in ER-negative breast cancer, but it is clear that these proteins represent promising targets for the treatment of ER-negative tumors.

References:

1. Ahr, A., et al., *Molecular classification of breast cancer patients by gene expression profiling.* J Pathol, 2001. **195**(3): p. 312-20.
2. Ivshina, A.V., et al., *Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer.* Cancer Res, 2006. **66**(21): p. 10292-301.
3. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
4. Sotiriou, C., et al., *Breast cancer classification and prognosis based on gene expression profiles from a population-based study.* Proc Natl Acad Sci U S A, 2003. **100**(18): p. 10393-8.
5. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.
6. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-6.
7. Chang, J.C., et al., *Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer.* Lancet, 2003. **362**(9381): p. 362-9.
8. Chang, J.C., et al., *Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients.* J Clin Oncol, 2005. **23**(6): p. 1169-77.
9. van 't Veer, L.J., et al., *Expression profiling predicts outcome in breast cancer.* Breast Cancer Res, 2003. **5**(1): p. 57-8.
10. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.
11. Finn, R.S., et al., *Dasatinib, an orally active small molecule inhibitor of both the src and abl kinases, selectively inhibits growth of basal-type/"triple-negative" breast cancer cell lines growing in vitro.* Breast Cancer Res Treat, 2007. **105**(3): p. 319-26.
12. Nishizuka, S., et al., *Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays.* Proc Natl Acad Sci U S A, 2003. **100**(24): p. 14229-34.
13. Chen, G., et al., *Protein profiles associated with survival in lung adenocarcinoma.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13537-42.
14. Neve, R.M., et al., *A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes.* Cancer Cell, 2006. **10**(6): p. 515-27.
15. Tyers, M. and M. Mann, *From genomics to proteomics.* Nature, 2003. **422**(6928): p. 193-7.

16. Richter, J., et al., *High-throughput tissue microarray analysis of cyclin E gene amplification and overexpression in urinary bladder cancer.* Am J Pathol, 2000. **157**(3): p. 787-94.
17. Schraml, P., et al., *Tissue microarrays for gene amplification surveys in many different tumor types.* Clin Cancer Res, 1999. **5**(8): p. 1966-75.
18. Bertucci, F., et al., *Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters.* Oncogene, 2004. **23**(7): p. 1377-91.
19. Wulfkuhle, J.D., et al., *New approaches to proteomic analysis of breast cancer.* Proteomics, 2001. **1**(10): p. 1205-15.
20. Hondermarck, H., et al., *Proteomics of breast cancer for marker discovery and signal pathway profiling.* Proteomics, 2001. **1**(10): p. 1216-32.
21. Bini, L., et al., *Protein expression profiles in human breast ductal carcinoma and histologically normal tissue.* Electrophoresis, 1997. **18**(15): p. 2832-41.
22. Kattenhorn, L.M., et al., *Identification of proteins associated with murine cytomegalovirus virions.* J Virol, 2004. **78**(20): p. 11187-97.
23. Petricoin, E.F., et al., *Use of proteomic patterns in serum to identify ovarian cancer.* Lancet, 2002. **359**(9306): p. 572-7.
24. Polley, A.C., et al., *Proteomic analysis reveals field-wide changes in protein expression in the morphologically normal mucosa of patients with colorectal neoplasia.* Cancer Res, 2006. **66**(13): p. 6553-62.
25. Sheehan, K.M., et al., *Use of reverse phase protein microarrays and reference standard development for molecular network analysis of metastatic ovarian carcinoma.* Mol Cell Proteomics, 2005. **4**(4): p. 346-55.
26. Tibes, R., et al., *Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells.* Mol Cancer Ther, 2006. **5**(10): p. 2512-21.
27. Nimeus, E., et al., *Proteomic analysis identifies candidate proteins associated with distant recurrences in breast cancer after adjuvant chemotherapy.* J Pharm Biomed Anal, 2007. **43**(3): p. 1086-93.
28. Stemke-Hale, K., et al., *Molecular screening for breast cancer prevention, early detection, and treatment planning: combining biomarkers from DNA, RNA, and protein.* Curr Oncol Rep, 2006. **8**(6): p. 484-91.
29. Carey, L.A., et al., *Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.* Jama, 2006. **295**(21): p. 2492-502.
30. Swaby, R.F., C.G. Sharma, and V.C. Jordan, *SERMs for the treatment and prevention of breast cancer.* Rev Endocr Metab Disord, 2007. **8**(3): p. 229-39.
31. Miller, K., et al., *Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer.* N Engl J Med, 2007. **357**(26): p. 2666-76.
32. Viani, G.A., et al., *Adjuvant trastuzumab in the treatment of her-2-positive early breast cancer: a meta-analysis of published randomized trials.* BMC Cancer, 2007. **7**: p. 153.

33. Dowell, J.E. and J.D. Minna, *EGFR mutations and molecularly targeted therapy: a new era in the treatment of lung cancer.* Nat Clin Pract Oncol, 2006. **3**(4): p. 170-1.
34. Herbst, R.S., M. Fukuoka, and J. Baselga, *Gefitinib--a novel targeted approach to treating cancer.* Nat Rev Cancer, 2004. **4**(12): p. 956-65.
35. Minna, J.D. and J. Dowell, *Erlotinib hydrochloride.* Nat Rev Drug Discov, 2005. **Suppl**: p. S14-5.
36. Dancey, J. and E.A. Sausville, *Issues and progress with protein kinase inhibitors for cancer treatment.* Nat Rev Drug Discov, 2003. **2**(4): p. 296-313.
37. Tan, A.R. and S.M. Swain, *Adjuvant chemotherapy for breast cancer: an update.* Semin Oncol, 2001. **28**(4): p. 359-76.
38. Hu, Z., et al., *The molecular portraits of breast tumors are conserved across microarray platforms.* BMC Genomics, 2006. **7**: p. 96.
39. Le Good, J.A., et al., *Protein kinase C isotypes controlled by phosphoinositide 3-kinase through the protein kinase PDK1.* Science, 1998. **281**(5385): p. 2042-5.
40. Alessi, D.R., et al., *3-Phosphoinositide-dependent protein kinase 1 (PDK1) phosphorylates and activates the p70 S6 kinase in vivo and in vitro.* Curr Biol, 1998. **8**(2): p. 69-81.
41. Kushner, P.J., et al., *Estrogen receptor pathways to AP-1.* J Steroid Biochem Mol Biol, 2000. **74**(5): p. 311-7.
42. Altucci, L., et al., *17beta-Estradiol induces cyclin D1 gene transcription, p36D1-p34cdk4 complex activation and p105Rb phosphorylation during mitogenic stimulation of G(1)-arrested human breast cancer cells.* Oncogene, 1996. **12**(11): p. 2315-24.
43. Geum, D., et al., *Estrogen-induced cyclin D1 and D3 gene expressions during mouse uterine cell proliferation in vivo: differential induction mechanism of cyclin D1 and D3.* Mol Reprod Dev, 1997. **46**(4): p. 450-8.
44. Dong, L., et al., *Mechanisms of transcriptional activation of bcl-2 gene expression by 17beta-estradiol in breast cancer cells.* J Biol Chem, 1999. **274**(45): p. 32099-107.
45. Zhang, X. and D. Yee, *Tyrosine kinase signalling in breast cancer: insulin-like growth factors and their receptors in breast cancer.* Breast Cancer Res, 2000. **2**(3): p. 170-5.
46. Kahlert, S., et al., *Estrogen receptor alpha rapidly activates the IGF-1 receptor pathway.* J Biol Chem, 2000. **275**(24): p. 18447-53.
47. Reis-Filho, J.S., et al., *FGFR1 emerges as a potential therapeutic target for lobular breast carcinomas.* Clin Cancer Res, 2006. **12**(22): p. 6652-62.
48. Elbauomy Elsheikh, S., et al., *FGFR1 amplification in breast carcinomas: a chromogenic in situ hybridisation analysis.* Breast Cancer Res, 2007. **9**(2): p. R23.

49. Bautista, S. and C. Theillet, *CCND1 and FGFR1 coamplification results in the colocalization of 11q13 and 8p12 sequences in breast tumor nuclei.* Genes Chromosomes Cancer, 1998. **22**(4): p. 268-77.
50. Johnston, S.R., et al., *Integration of signal transduction inhibitors with endocrine therapy: an approach to overcoming hormone resistance in breast cancer.* Clin Cancer Res, 2003. **9**(1 Pt 2): p. 524S-32S.
51. Massarweh, S. and R. Schiff, *Unraveling the mechanisms of endocrine resistance in breast cancer: new therapeutic opportunities.* Clin Cancer Res, 2007. **13**(7): p. 1950-4.
52. Massarweh, S. and R. Schiff, *Resistance to endocrine therapy in breast cancer: exploiting estrogen receptor/growth factor signaling crosstalk.* Endocr Relat Cancer, 2006. **13 Suppl 1**: p. S15-24.
53. Varambally, S., et al., *Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression.* Cancer Cell, 2005. **8**(5): p. 393-406.
54. Tian, Q., et al., *Integrated genomic and proteomic analyses of gene expression in Mammalian cells.* Mol Cell Proteomics, 2004. **3**(10): p. 960-9.

Supplementary Table 4.1- List of protein antibodies used in the reverse phase protein lysate arrays.

Supplemental Table 4.1. Eighty-nine monospecific antibodies used in this study.

Antibody name	Protein name	Company	cat#	Host	Dilution
4EBP1	4E Binding Protein 1	Cell Signaling Technology, Inc.	CS 9452	Rabbit	1 in 100
4EBP1p37	4EBP1 phosphorylation at T37/T46	Cell Signaling Technology, Inc.	CS 9459	Rabbit	1 in 100
AcCoA	Acetyl CoA Carboxylase	Epitomics, Inc.	1768-1	Rabbit	1 in 250
AcCoAp	AcCoA phosphorylation at S79	Cell Signaling Technology, Inc.	CS 3661	Rabbit	1 in 250
Akt	Protein Kinase B	Cell Signaling Technology, Inc.	CS 9272	Rabbit	1 in 250
Aktp308	Akt phosphorylation at S308	Cell Signaling Technology, Inc.	CS 9275	Rabbit	1 in 250
Aktp473	Akt phosphorylation at S473	Cell Signaling Technology, Inc.	CS 9271	Rabbit	1 in 250
AMPK	AMPK	Cell Signaling Technology, Inc.	CS 2532	Rabbit	1 in 250
AMPKp	AMPK phosphorylation at S172	Cell Signaling Technology, Inc.	CS 2535	Rabbit	1 in 250
AR	Androgen receptor	Epitomics	1852-1	Rabbit	1 in 200
β catenin	B catenin	Cell Signaling Technology, Inc.	CS 9562	Rabbit	1 in 300
bcl2	bcl2	Dako	M0887	Mouse	1 in 200
BRCA1	BRCA1	Upstate Biotechnology, Inc.	07-434	Rabbit	1 in 1000
caveolin 1	Caveolin 1	Cell Signaling Technology, Inc.	CS 3232	Rabbit	1 in 250
CCNB1	Cyclin B1	Epitomics, Inc.	1495-1	Rabbit	1 in 500
CCND1	Cyclin D1	Santa Cruz Biotechnology, Inc.	SC-718	Rabbit	1 in 1000
CCNE1	Cyclin E1	Santa Cruz Biotechnology, Inc.	SC-247	Mouse	1 in 500
CCNE2	Cyclin E2	Epitomics	1142-1	Rabbit	1 in 250
CD31	CD31	Dako	M0823	Mouse	1 in 500
CDK4	CDK4	Cell Signaling Technology, Inc.	CS 2906	Rabbit	1 in 250
cjun	Cjun	Cell Signaling Technology, Inc.	CS 9165	Rabbit	1 in 250
ckit	Ckit	Cell Signaling Technology, Inc.		Rabbit	1 in 150
cleaved caspase 7	Cleaved caspase 7 (Asp198)	Cell Signaling Technology, Inc.	CS 9491	Rabbit	1 in 150
cleaved PARP	Cleaved PARP (Asp214)	Cell Signaling Technology, Inc.	CS 9546	Mouse	1 in 250
cmyc	Cmyc	Cell Signaling Technology, Inc.	CS 9402	Rabbit	1 in 150
Collagen VI	Collagen VI	Santa Cruz Biotechnology, Inc.	SC-20649	Rabbit	1 in 750
COX2	COX2	Epitomics, Inc.	2169-1	Rabbit	1 in 500
E cadherin	E cadherin	Cell Signaling Technology, Inc.	CS 4065	Rabbit	1 in 200
EGFR	Epidermal growth factor receptor	Santa Cruz Biotechnology, Inc.	SC-03	Rabbit	1 in 200
EGFRp1045	EGFR phosphorylation at Y1045	Cell Signaling Technology, Inc.	CS 2237	Rabbit	1 in 100

EGFRp922	EGFR phosphorylation at Y992	Cell Signaling Technology, Inc.	CS 2235	Rabbit	1 in 100
ER	Estrogen receptor alpha	Lab Vision Coorporation (formerly Neomarkers)	Sp1	Rabbit	1 in 250
ERK2	Mitogen-activated protein kinase	Cell Signaling Technology, Inc.	SC-154	Rabbit	1 in 250
ERp118	ER phosphorylation at S118	Epitomics, Inc.	1091-1	Rabbit	1 in 200
ERp167	ER phosphorylation at S167	Epitomics, Inc.	2492-1	Rabbit	1 in 200
FGFR1	FGFR1	Santa Cruz	SC-7945	Rabbit	1 in 250
GATA3	GATA binding protein 3	BD Biosciences	558686	Mouse	1 in 200
GSK3	Glycogen synthase kinase 3 beta	Santa Cruz Biotechnology, Inc.	SC-7291	Mouse	1 in 1000
GSK3p21_9	GSK3 phosphorylation at S21/S9	Cell Signaling Technology, Inc.	CS 9331	Rabbit	1 in 250
HER2	Human epidermal receptor 2	Epitomics, Inc.	1148-1	Rabbit	1 in 250
HER2p1248	HER2 phosphorylation at Y1248	Upstate Biotechnology, Inc.	06-229	Rabbit	1 in 750
IGF1R	Insulin-like growth factor receptor 1	Cell Signaling Technology, Inc.	CS 3027	Rabbit	1 in 500
IGFRp	IGF1R phosphorylation at Y1135/Y1136	Cell Signaling Technology, Inc.	CS 3024	Rabbit	1 in 200
JNK	cjun N terminal Kinase	Santa Cruz Biotechnology, Inc.	SC-474	Rabbit	1 in 200
JNKp183-185	JNK phosphorylation at T183/Y185	Cell Signaling Technology, Inc.	CS 9251	Rabbit	1 in 150
LKB1	LKB1	Abcam	15095	Mouse	1 in 200
LKB1p	LKB1p	Cell Signaling Technology, Inc	SC-3054	Rabbit	1 in 250
MAPKp	MAPK1/2 phosphorylation at T202/T204	Cell Signaling Technology, Inc.	CS 4377	Rabbit	1 in 1000
MEK1	MAPK/ERK kinase 1	Epitomics, Inc.	1235-1	Rabbit	1 in 15000
MEK12p	MEK1/2 phosphorylation at T217/T221	Cell Signaling Technology, Inc.	CS 9121	Rabbit	1 in 800
mTOR	mammalian target of rapamycin	Cell Signaling Technology, Inc.	CS 2983	Rabbit	1 in 400
p110alpha	p110alpha subunit of phosphatidylinositol-3-kinase	Epitomics, Inc.	1683-1	Rabbit	1 in 500
p21	p21	Santa Cruz Biotechnology, Inc.	SC-397	Rabbit	1 in 250
p27	p27	Santa Cruz Biotechnology, Inc.	SC-527	Rabbit	1 in 500
p38	p38 MAPK	Cell Signaling Technology, Inc.	CS 9212	Rabbit	1 in 300
p38p180_2	p38 MAPK phosphorylation at T180/T182	Cell Signaling Technology, Inc.	CS 9211	Rabbit	1 in 250
p53	p53	Cell Signaling Technology, Inc.	CS 9282	Rabbit	1 in 3000
p7056 Kinase	p70S6 Kinase	Epitomics, Inc.	1494-1	Rabbit	1 in 500
p70S6Kp389	p70S6 Kinase phosphorylation at T389	Cell Signaling Technology, Inc.	CS 9205	Rabbit	1 in 200
PAI1	Plasminogen activator inhibitor-1	BD Biosciences	612024	Mouse	1 in 1000
pcmyc	cmyc phosphorylation at T58/S62	Cell Signaling Technology, Inc.	CS 9401	Rabbit	1 in 150

PDK1	Phosphoinositide Dependent Kinase 1	Cell Signaling Technology, Inc.	CS 3062	Rabbit	1 in 250
PDK1p241	PDK1 phosphorylation at S241	Cell Signaling Technology, Inc.	CS 3061	Rabbit	1 in 500
PKCalpha	Protein Kinase C alpha	Upstate Biotechnology, Inc.	05-154	Mouse	1 in 2000
PKCaphap657	PKCalpha phosphorylation at S657	Upstate Biotechnology, Inc.	06-822	Rabbit	1 in 3000
pmTOR	mTOR phosphorylation at S2448	Cell Signaling Technology, Inc.	CS 2971	Rabbit	1 in 150
PR	Progesterone receptor	Epitomics, Inc.	1483-1	Rabbit	1 in 400
PTEN	PTEN	Cell Signaling Technology, Inc.	CS 9552	Rabbit	1 in 500
Rab25	Rab25	Courtesy Dr. Kwai Wa Cheng, MDACC	Covance	Rabbit	1 in 4000
Rb	Retinoblastoma	Cell Signaling Technology, Inc.	CS 9309	Mouse	1 in 3000
Rbp	Rb phosphorylation at S807/S811	Cell Signaling Technology, Inc.	CS 9308	Rabbit	1 in 250
S6	S6 ribosomal protein	Cell Signaling Technology, Inc.	CS 2217	Rabbit	1 in 200
S6p235-236	S6 phosphorylation at S235/S236	Cell Signaling Technology, Inc.	CS 2211	Rabbit	1 in 3000
S6p240_4	S6 phosphorylation at S240/S244	Cell Signaling Technology, Inc.	CS 2215	Rabbit	1 in 3000
SGK	Serum Glucocorticoid Kinase	Cell Signaling Technology, Inc.	CS 3272	Rabbit	1 in 250
SGKp	SGK phosphorylation at S78	Cell Signaling Technology, Inc.	CS 3271	Rabbit	1 in 250
src	Src	Upstate Biotechnology, Inc.	05-184	Mouse	1 in 200
srcp416	src phosphorylation at Y416	Cell Signaling Technology, Inc.	CS 2101	Rabbit	1 in 150
srcp527	src phosphorylation at Y527	Cell Signaling Technology, Inc.	CS 2105	Rabbit	1 in 400
stat3	STAT3	Upstate Biotechnology, Inc.	06-596	Rabbit	1 in 500
stat3p705	stat3 phosphorylation at S705	Cell Signaling Technology, Inc.	CS 9131	Rabbit	1 in 500
stat3p727	stat3 phosphorylation at S727	Cell Signaling Technology, Inc.	CS 9134	Rabbit	1 in 250
stat6p641	stat6 phosphorylation at Y641	Cell Signaling Technology, Inc.	CS 9361	Rabbit	1 in 150
stathmin	Stathmin	Epitomics, Inc.	1972-1	Rabbit	1 in 500
TSC2	Tuberous Sclerosis Kinase 2	Epitomics, Inc.	1613-1	Rabbit	1 in 500
TSC2p	TSC2 phosphorylation at T1462	Cell Signaling Technology, Inc.	CS 3617	Rabbit	1 in 200
VEGFR2	KDR2 / VEGF Receptor 2	Cell Signaling Technology, Inc.	CS 2479	Rabbit	1 in 700
XIAP	X linked inhibitor of apoptosis	Cell Signaling Technology, Inc.	CS 2042	Rabbit	1 in 200

Companies
Abcam, Inc. (Cambridge, MA), BD Biosciences (San Jose, CA), Cell Signaling Technology, Inc. (Danvers, MA), Dako (Carpinteria, CA), Epitomics, Inc. (Burlingame, CA), Santa Cruz Biotechnology, Inc. (Santa Cruz, CA), Upstate Biotechnology, Inc. (Millipore)

Chapter 5

Maternal Embryonic Leucine Zipper Kinase (MELK) is a Key Regulator of Proliferation and is Independently Prognostic in Estrogen Receptor-Negative Breast Cancer

5.1 Abstract

Cancer therapies directed at specific molecular targets in signaling pathways of cancer cells, such as tamoxifen, aromatase inhibitors and trastuzumab, have proven useful for treatment of breast cancer. However, these targeted therapies have significant side effects and do not effectively treat or cure all breast cancers, especially estrogen receptor alpha (ER)-negative, progesterone receptor (PR)-negative, HER2-negative ("triple-negative") breast cancer. The long term goal of our study is to identify critical growth regulatory molecules in ER-negative breast cancer that could be targeted for the treatment of these aggressive breast cancers. Using gene expression profiling of human breast cancers, we discovered that maternal embryonic leucine-zipper kinase (MELK) is significantly overexpressed in ER-negative breast cancers as compared to ER-positive breast cancers. We demonstrated that MELK was overexpressed in ER-negative breast cancers in independent breast tumor data sets using gene expression profiling and quantitative real time polymerase chain reaction (Q-RT-PCR) analysis. MELK was also significantly more highly expressed at the RNA and protein levels in ER-negative breast cancer cell lines as compared to ER-positive breast cancer cell lines. We next conducted MELK expression knockdown studies which demonstrated that MELK is essential for the growth of most ER-negative, but not ER-positive, breast cancer cell lines. Finally, survival analysis of patients with breast cancer shows that those patients whose tumors have high expression of MELK have a significantly poorer prognosis than patients with low expression of MELK. In multivariate analysis, MELK is an independent prognostic

factor in breast cancer. These results show that MELK is highly expressed in ER-negative, including “triple-negative” breast cancer, and that it is essential for breast cancer cell growth. This study suggests that MELK is a promising target for the treatment of ER-negative, and especially “triple-negative”, breast cancer.

5.2 Introduction

Estrogen receptor alpha (hereafter referred to as ER)-positive breast cancers account for 60-70% of breast cancers, while the remaining 30-40% of breast cancers are ER-negative and poorly responsive to traditional therapies [1]. Selective estrogen receptor modulators (SERMs), such as tamoxifen and raloxifene, and aromatase inhibitors have been shown to reduce ER-positive breast cancer recurrence by approximately 50% [2]. These agents, however, are not effective for the treatment of ER-negative breast cancer. Currently, chemotherapy is used to treat ER-negative tumors [3]. Such therapy is generally toxic and is not specifically targeted to ER-negative breast cancer, instead only non-specifically killing rapidly dividing cells.

A major goal of current breast cancer research has been to identify targets that are unique to cancer cells which are critical for the growth and survival of the cancer cells. While achieving this goal has been difficult, there are several examples of effective targeted therapies, including the monoclonal antibodies trastuzumab (targeting the HER2/*neu* receptor) and bevacizumab (targeting vascular epithelial growth factor) [4, 5]. Other targeted therapies include the small molecule tyrosine kinase inhibitors gefitinib and erlotinib (both of which target the epidermal growth factor receptor), and lapatinib (a dual kinase inhibitor targeting both the epidermal growth factor receptor and the HER2/*neu* receptor) [6-9]. However, these therapies are effective only in approximately 20% of patients whose tumors overexpress HER2. Additionally, due to deleterious side effects of these drugs and the emergence of

therapeutic drug resistance, it is necessary to discover additional critical growth regulatory molecules that can be targeted for the treatment of ER-negative breast cancer.

To identify novel targets for the treatment of ER-negative breast cancers, including the aggressive ER-negative, PR-negative, HER2-negative ("triple-negative") breast cancers, we used expression microarray analysis to identify molecules that are highly expressed in ER-negative breast cancer that may play a role in breast cancer development and progression. Through these studies we identified maternal embryonic leucine zipper kinases (MELK) as a potential target for the treatment of ER-negative breast cancer. MELK was identified as being significantly overexpressed in ER-negative breast tumors.

We demonstrate that MELK expression is significantly elevated in multiple breast cancer datasets as well as in ER-negative breast cancer cell lines. We also show that MELK expression is critical for the growth of ER-negative, but not ER-positive, breast cancer cells. Furthermore, we show that MELK expression is significantly associated with poor metastasis-free and overall survival, and that it is independently prognostic in breast cancer. These results implicate MELK as a promising target for the treatment of ER-negative breast cancer.

5.3 Results

To identify novel targets for the treatment of ER-negative breast cancers, we first identified kinases that were overexpressed in ER-negative breast tumors using Affymetrix gene expression profiling.

5.3.1 Patient Population

For the profiling experiments, the same tumors and patient information described in **Chapter 3** were used, and are again described here. A total of 102 patients with invasive breast cancer were recruited through IRB-approved, neoadjuvant clinical trials to investigate gene expression in human tumors before and after drug treatment by Dr. Jenny Chang at Baylor College of Medicine. Breast biopsies taken before initiation of any treatment were used in this study. Because the patients did not receive systemic adjuvant or neoadjuvant therapy prior to the biopsy, the results from the gene expression analysis represent basal gene expression in these breast cancers. For the gene expression profiling experiments, 102 breast tumors were studied, 58 of which were ER-positive and 44 ER-negative by IHC-staining (24 of which were confirmed as "triple-negative"). The tumors were all stage III or IV from pre- and post-menopausal women, with all tumors showing >30% cellularity. The women were from several racial groups and the majority had no palpable nodes at baseline. Most of the women were premenopausal and presented with relatively large tumors (ranging from 2.5 to 25 cm). The clinical and demographic features of these tumors are summarized in **Table 5.1**.

Table 5.1. - **Clinical characteristics of the patients and tumor samples used to identify MELK overexpression by Affymetrix gene expression profiling.**

Characteristic	Tumor Set N=102 (%)
Age	
Mean	48.1
Range	(32-72)
Race	
Caucasian	50 (57%)
Hispanic	7 (8%)
African-American	23 (27%)
Asian	7 (8%)
Menopausal Status	
Pre	49 (62%)
Post	30 (38%)
BMI	
Mean	29.7
Range	(16.1-48.3)
Baseline Tumor Size, cm	
Mean	6.3
Range	(2.5-25.0)
Palpable Nodes at Baseline	
Yes	20 (21%)
No	77 (79%)
ER	
Positive	57 (56%)
Negative	45 (44%)
Unknown	0 (0%)
PR	
Positive	37 (36%)
Negative	47 (46%)
Unknown	18 (18%)
HER2/Neu	
Positive	27 (26%)
Negative	58 (57%)
Unknown	17 (17%)

Table 5.1- Characteristics of 102 patients with breast cancer. Tumors from these patients were used for gene expression profiling to identify MELK as being overexpressed in ER-negative breast tumors. The tumors were collected and profiled by Dr. Jenny Chang at Baylor College of Medicine.

These tumors were obtained, processed, and profiled in the laboratory of Dr. Jenny Chang. All analyses in this chapter were done by Corey Speers with input from Dr. Susan Hilsenbeck at Baylor College of Medicine, unless otherwise indicated.

5.3.2 Affymetrix Gene Expression Profiling Identified Maternal Embryonic Leucine Zipper Kinase (MELK) as One of the Most Overexpressed Kinases in Human ER-Negative Breast Tumors

This gene expression profiling study identified 52 kinases that were overexpressed in ER-negative breast cancer compared to ER-positive breast cancers (*P*-value < 0.05) (a list of the 52 kinases is shown in **Table 3.4**). The Affymetrix techniques and statistical analysis are described in **Chapter 2** and were done in the laboratory of Dr. Jenny Chang**.** One of the most highly expressed kinase in ER-negative breast cancer was maternal embryonic leucine zipper kinase (MELK), which was expressed 3.75 fold higher in the ER-negative tumor samples. A list of the 10 most highly overexpressed kinases in this analysis is found in **Table 5.2**. Because little was known about the function of MELK, or its potential role in carcinogenesis, we determined its expression level in ER-positive and ER-negative breast cancers using other datasets. These datasets and their characteristics are described in **Chapter 2.2.7**.

Table 5.2 Top 10 most highly expressed in ER-negative breast cancer

Fold difference (ER-/ER+)	Description	Gene symbol	Probe set	Permutation p-value
5.49	phosphofructokinase, platelet	PFKP	201037_at	< 0.001
5.42	CHK1 checkpoint homolog (S. pombe)	CHEK1	205393_s_at	< 0.001
4.18	uridine-cytidine kinase 2	UCK2	209825_s_at	< 0.001
3.81	chemokine (C-X-C motif) ligand 10	CXCL10	204533_at	0.011
3.75	maternal embryonic leucine zipper kinase	MELK	204825_at	0.009
3.72	v-yes-1 viral oncogene homolog 1	YES1	202932_at	0.032
3.71	cell division cycle 2, G1 to S and G2 to M	CDC2	210559_s_at	< 0.001
3.42	pyridoxal (pyridoxine, vitamin B6) kinase	PDXK	202671_s_at	< 0.001
3.40	mitogen-activated protein kinase kinase kinase 5	MAP3K5	203836_s_at	0.01
3.40	BUB1 homolog (yeast)	BUB1	209642_at	0.003

Table 5.2 - List of the 10 kinases and kinase associated proteins with highest expression in ER-negative tumors compared to ER-positive tumors.

5.3.3 Confirmation of High MELK Expression in an Independent Set of Human Breast Tumors using Q-RT-PCR

We also wanted to confirm that MELK is more highly expressed in ER-negative than in ER-positive breast cancers using another technique. We therefore used an independent set of 60 human breast tumors obtained from the tumor bank at Baylor College of Medicine for further validation. After identifying equal numbers of ER-positive and negative samples (30 of each), we used quantitative RT-PCR (Q-RT-PCR) to confirm higher expression of MELK in ER-negative tumors. Q-RT-PCR analysis showed that MELK expression was significantly higher in ER-negative breast tumor samples compared to ER-positive tumors (*P*-value 0.02) (**Figure 5.1**). This set of data validated the gene expression profiling experiments and confirmed that MELK is indeed more highly expressed in ER-negative breast cancer.

5.3.4 SAGE Analysis Identifies MELK as being Highly Expressed in Cancerous, but not Normal, Tissue

To gain insight into the role of MELK in ER-negative breast carcinogenesis, we analyzed serial analysis of gene expression (SAGE) data that was drawn from multiple human tissues and cancer cell lines. This analysis demonstrated that MELK was overexpressed at a significantly higher level in a great majority of cancerous tissues, especially ovary, breast, and prostate, but was not expressed at appreciable levels in normal vital organs including heart, liver, lung and kidney, among others (**Figure 5.2**).

Figure 5.1 – Q-RT-PCR analysis of RNA expression from an independent set of human breast tumors shows MELK is significantly more highly expressed in ER-negative breast tumors. The expression of MELK in 60 human breast tumor samples (30 ER-negative and 30 ER-positive) was measured using Q-RT-PCR analysis. Levels of MELK expression were confirmed as being significantly higher in ER-negative human breast tumors than in ER-positive human breast tumors. Data are represented as mean ± SD. Experiment and analysis done by Corey Speers.

Figure 5.1

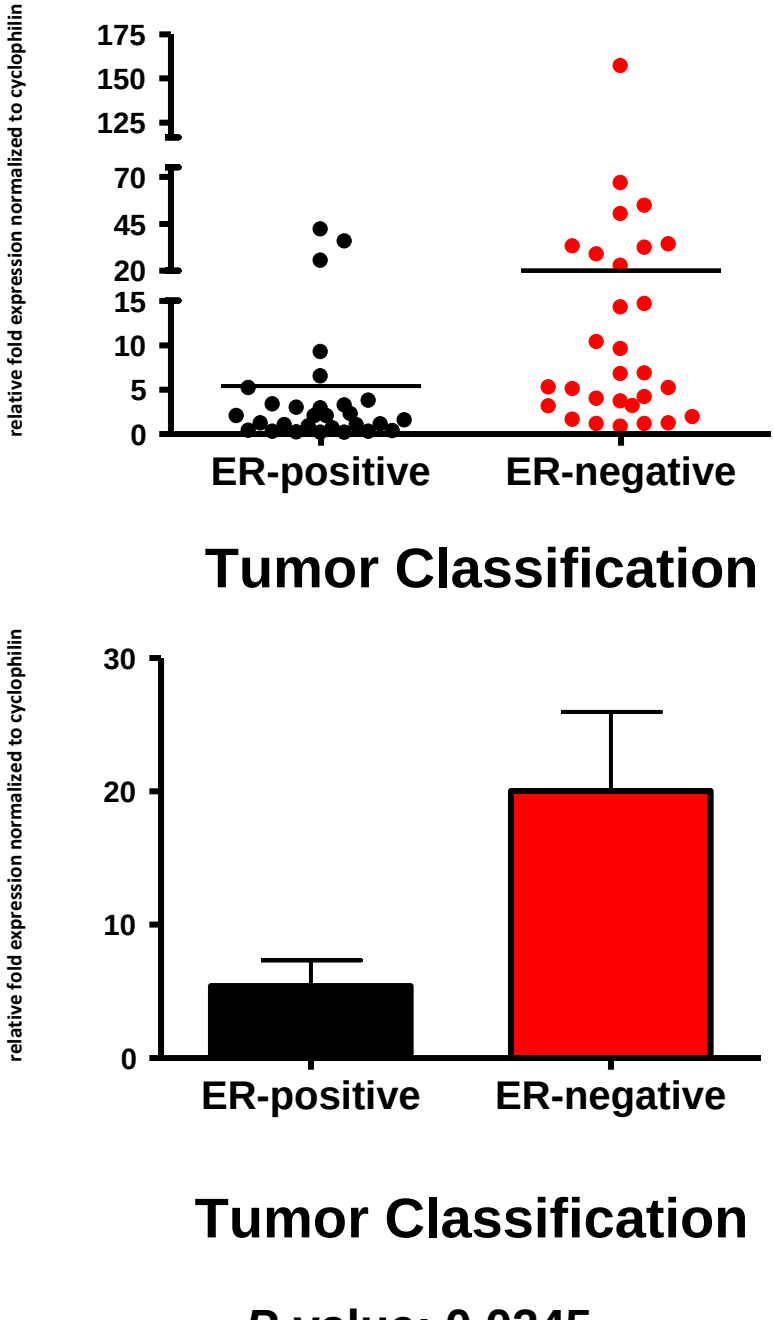

relative fold expression normalized to cyclophilin
175
150
125
70
45
20
15
10
5
0
ER-positive
ER-negative
Tumor Classification
relative fold expression normalized to cyclophilin
30
20
10
0
ER-positive
ER-negative
Tumor Classification
P-value: 0.0245

Figure 5.2 SAGE Analysis Comparing Normal to Cancer

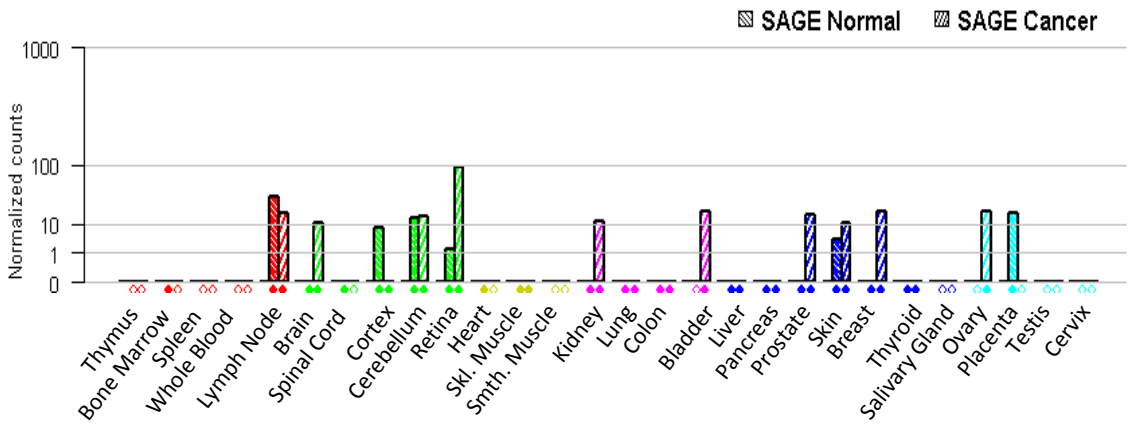


Figure 5.2 – Serial analysis of gene expression (SAGE) analysis in the NCI-160 shows that MELK is more highly expressed in cancerous tissue, especially when compared to normal tissue. Prostate, Kidney, Ovary, and Breast tissues have little to no expression of MELK in normal tissues, but expression is increased at least 10-fold in cancerous tissue. Analysis done by Corey Speers.

5.3.5 Differentially Expressed Kinases Validated Using Publicly Available Data Sets

To demonstrate that MELK was indeed more highly expressed in ER-negative tumors compared to ER-positive tumors, we analyzed gene expression data from 12 additional publically available data sets as a validation method. These datasets from multiple investigators include over 1800 additional tumor samples (556 ER-negative and 1282 ER-positive tumors) for which there is gene expression profiling data and constitutes the most comprehensive breast tumor set currently available [10-20]. To utilize the power of such a large combined dataset, we employed a technique recently described by Whitlock that relies on a weighted Z-method to combine *P*-values [21]. This robust approach, superior to Fisher's combined probability test, revealed that MELK was significantly more highly expressed (z-score of 19.76 and *P*-value < 1.0 e-20) in ER-negative breast tumors as compared to ER-positive tumors in an effective sample size of over 1800 tumors (**Table 5.3**). Many other kinases that are more highly expressed in ER-negative breast cancer are also listed in Table 5.3. These kinases represent additional targets of interest in the lab, and some are already the target of directed therapies (*EGFR, CHEK1, LYN, LCK, AURKB*). Many of these kinases are involved in cell cycle checkpoint regulation (*CHEK1. BUB1, TTK, CDC7, CDC2*), are involved in cellular metabolism (*PFKP, UCK2, PDXK, SEPHS1, PGK1, AK2, UGP2*) or are members of the Ephrin family of receptors and kinases (*EPHA2, EPHB6, EPHB2, EPHB4*). Because very little had been published regarding the function of MELK and its very high significance in

Table 5.3 - Kinases that are more highly expressed in ER-negative breast cancers

Gene symbol	Z-score	*P* -value	Gene symbol	Z-score	*P* -value
SRPK1	20.84	< 1.0 e-20	**AK2**	14.97	< 1.0 e-20
PRKX	20.39	< 1.0 e-20	**UGP2**	14.76	< 1.0 e-20
MELK	19.76	< 1.0 e-20	**LIMK2**	14.52	< 1.0 e-20
SRPK1	19.72	< 1.0 e-20	**PIM1**	14.44	< 1.0 e-20
PFKP	19.69	< 1.0 e-20	**CDC2**	14.36	< 1.0 e-20
EGFR	19.59	< 1.0 e-20	**LCK**	14.14	< 1.0 e-20
LYN	19.57	< 1.0 e-20	**MET**	13.80	< 1.0 e-20
CHEK1	19.02	< 1.0 e-20	**EPHA2**	13.73	< 1.0 e-20
BUB1	18.64	< 1.0 e-20	**CCL2**	13.59	< 1.0 e-20
DAPK1	18.50	< 1.0 e-20	**AURKB**	13.49	< 1.0 e-20
TTK	18.14	< 1.0 e-20	**RPS6KA3**	13.25	< 1.0 e-20
YES1	17.86	< 1.0 e-20	**EPHB6**	13.03	< 1.0 e-20
MAP4K4	17.81	< 1.0 e-20	**MAPK1**	12.35	< 1.0 e-20
STK38	17.81	< 1.0 e-20	**MAP3K5**	12.30	< 1.0 e-20
RIPK4	17.43	< 1.0 e-20	**EPHB2**	12.20	< 1.0 e-20
UCK2	17.02	< 1.0 e-20	**EPHB4**	12.19	< 1.0 e-20
CXCL10	16.87	< 1.0 e-20	**CSK**	12.18	< 1.0 e-20
PDXK	16.87	< 1.0 e-20	**CCL4**	10.86	< 1.0 e-20
SEPHS1	16.55	< 1.0 e-20	**STK38L**	10.80	< 1.0 e-20
IRAK1	16.51	< 1.0 e-20	**SGK**	10.40	< 1.0 e-20
MALT1	16.43	< 1.0 e-20	**RIOK3**	9.93	2.2 e-20
PLK1	15.89	< 1.0 e-20	**VRK2**	8.90	1.0 e-17
YWHAQ	15.46	< 1.0 e-20	**RPS6KA1**	5.51	0.000000005
PGK1	15.20	< 1.0 e-20	**MAP4K2**	5.15	0.00000008
CDC7	15.19	< 1.0 e-20	**RYK**	4.55	0.0000008
PTK7	15.16	< 1.0 e-20	**PIK3CB**	4.31	0.000009

Table 5.3- High MELK expression validated in independent human tumor sample data sets.
Data analysis of an additional 12 publically available human breast tumor datasets shows MELK expression levels as being more highly expressed in ER-negative breast tumors in the other datasets. Z-scores were calculated using the Z-transform test and are listed with their correlating *P* -value.

multiple datasets being more highly expressed in ER-negative disease, we chose to focus initial studies on MELK.

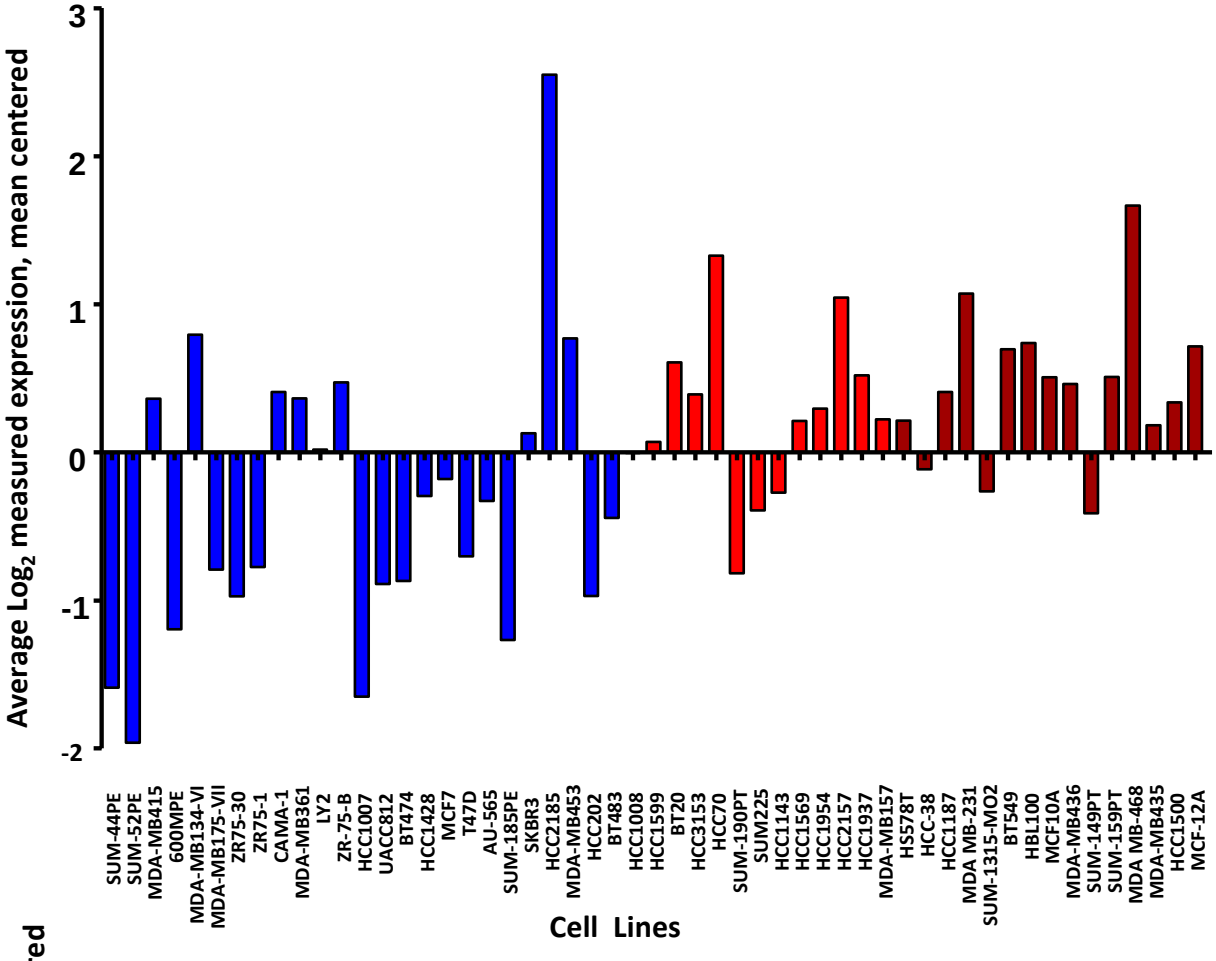
5.3.6 MELK is More Highly Expressed in ER-Negative Breast Cancer Cell Lines

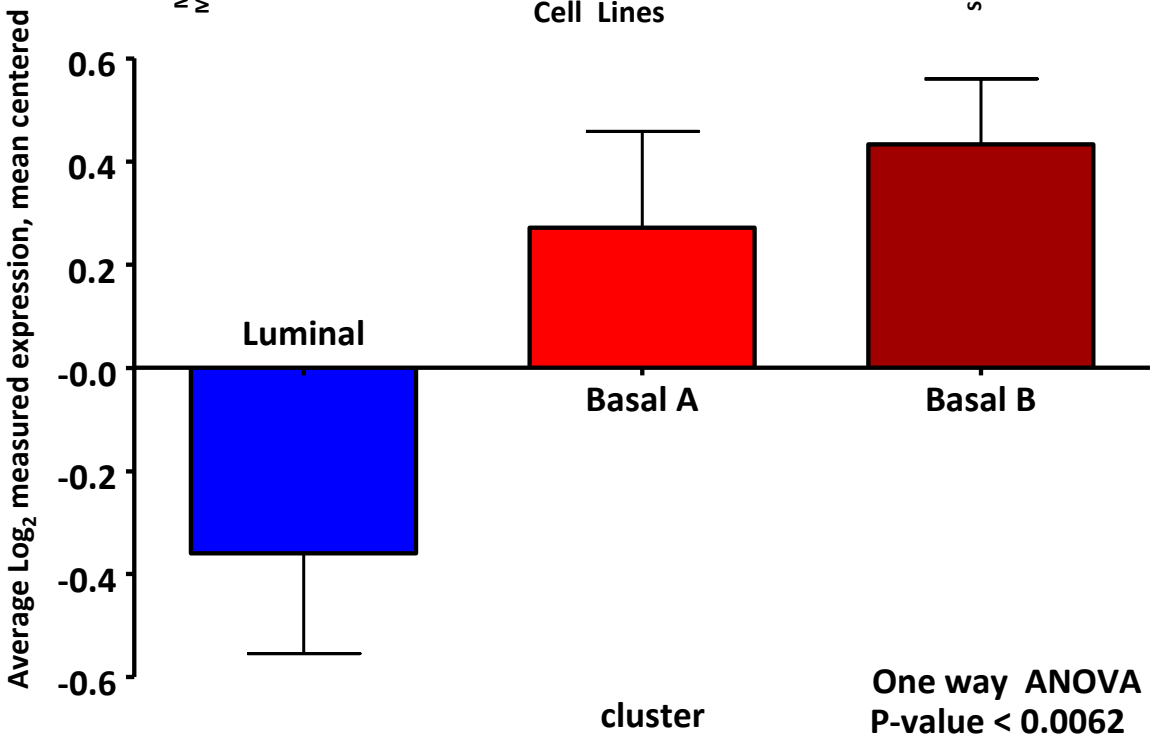
To conduct further *in vitro* experimentation in cell lines, we needed to determine the MELK expression levels in a number of breast cancer cell lines. Recent work by Neve *et al*. showed that the recurrent genomic and transcriptional characteristics of breast cancer cell lines mirror those of primary breast tumors (28). These investigators performed Affymetrix gene expression profiling on a set of 51 commonly used ER-positive and ER-negative breast cancer cell lines and used hierarchical clustering to show that the cell lines clustered into three main groups: basal A, basal B, and luminal (28). We used this expression information from breast cancer cell lines to determine the expression level of MELK in these 51 breast cancer cell lines. This data shows that MELK is expressed more highly in the ER-negative (Basal A and Basal B) breast cancer cell lines (**Figure 5.3**). It also showed that the ER-negative cell lines HCC 2185, HCC 70, MDA-MB-231, MDA-MB-468, and HCC 1937 have very high levels of MELK expression, while the ER-positive cell lines SUM-44PE, SUM-52PE, HCC 1007, and SUM-185PE have very low expression of MELK.

As this gene expression data indicated that MELK was overexpressed in ER-negative breast cancer cell lines, we chose twelve ER-positive or ER-negative breast cancer cell lines and measured the expression of MELK under basal growth conditions

Figure 5.3 – Analysis of gene expression data from 51 breast cancer cell lines shows MELK is significantly overexpressed in ER-negative (Basal A and Basal B) cell lines. Mean centered log2 transformed gene expression values from 51 common breast cancer cell lines shows that MELK expression is significantly higher in ER-negative breast cancer cell lines, including cell lines that are Basal A and Basal B as defined by Neve *et al*. Cell lines including HCC 70, HCC 1954, MDA-MB-231, and MDA-MB-468 have especially high levels of MELK expression. Data are represented as mean ± SD. Red color indicate ER-negative breast cancer cell lines, green color indicated ER-positive cell lines. Data was obtained from Dr. Joe Gray and analysis was done by Corey Speers.

Figure 5.3


Average Log2 measured expression, mean centered
3
2
1
0
-1
-2
SUM-44PE
SUM-52PE
MDA-MB415
600MPE
MDA-MB134-VI
MDA-MB175-VII
ZR75-30
ZR75-1
CAMA-1
MDA-MB361
LY2
ZR-75-B
HCC1007
UACC812
BT474
HCC1428
MCF7
T47D
AU-565
SUM-185PE
SKBR3
HCC2185
MDA-MB453
HCC202
BT483
HCC1008
HCC1599
BT20
HCC3153
HCC70
SUM-190PT
SUM225
HCC1143
HCC1569
HCC1954
HCC2157
HCC1937
MDA-MB157
HS578T
HCC-38
HCC1187
MDA MB-231
SUM-1315-MO2
BT549
HBL100
MCF10A
MDA-MB436
SUM-149PT
SUM-159PT
MDA MB-468
MDA-MB435
HCC1500
MCF-12A
Cell Lines


Average Log2 measured expression, mean centered
0.6
0.4
0.2
-0.0
-0.2
-0.4
-0.6
Luminal
Basal A
Basal B
cluster
One way ANOVA
P-value < 0.0062

using Q-RT-PCR. MELK expression was again found to be significantly elevated in the ER-negative breast cancer cell lines (P-value 0.007) as compared to the ER-positive breast cancer cell lines (**Figure 5.4**).

We next verified that the protein levels of MELK were also increased in ER-negative breast cancer cell lines. Using western blot analysis, we demonstrated that MELK was more highly expressed in ER-negative breast cancer cell lines, thus correlating with the RNA expression information obtained with gene expression profiling and Q-RT-PCR (**Figure 5.5**). This RNA and protein expression data was used to design experiments examining the effect of overexpressing MELK in cell lines with normally low endogenous expression (i.e. SUM-44PE) or knockdown studies in cell lines that have high expression of MELK (i.e. MDA-MB-468).

5.3.7 Effect of MELK Knockdown on the Growth of ER-negative and ER-positive Breast Cancer Cell Lines

We next investigated whether MELK was critical for the growth of ER-negative breast cancer. Using siRNA knockdown of MELK RNA levels, we were able to determine the effect of MELK inhibition on breast cancer cell proliferation in 9 breast cancer cell lines. ER-positive (MCF-7, MDA-MB-361, and T47D) and ER-negative (MDA-MB-468, MDA-MB-231, HCC 1937, HCC 1187, HCC 1569, BT549, and Hs578T) cells were transfected with siRNA designed to knockdown the expression of MELK. The siRNA construct used in this study showed at least 70% knockdown of target kinase expression

Figure 5.4 – Q-RT-PCR analysis of RNA expression from 14 breast cancer cell lines shows MELK is significantly more highly expressed in ER-negative breast cell lines. The expression of MELK in 14 breast cancer cell lines (9 ER-negative and 5 ER-positive) was measured using Q-RT-PCR analysis. Data is depicted as relative fold change normalized to cyclophilin. Data are represented as mean ± SD. Red color indicates ER-negative cell lines, green indicates ER-positive cell lines. Experimentation and analysis done by Corey Speers.

Figure 5.4

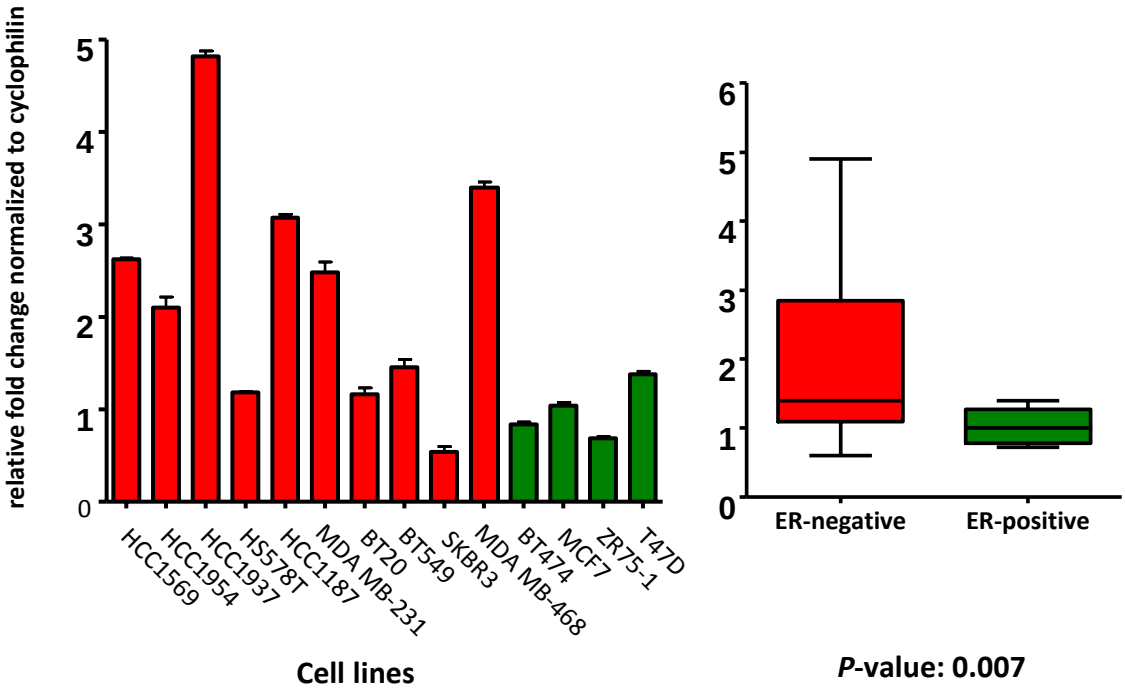
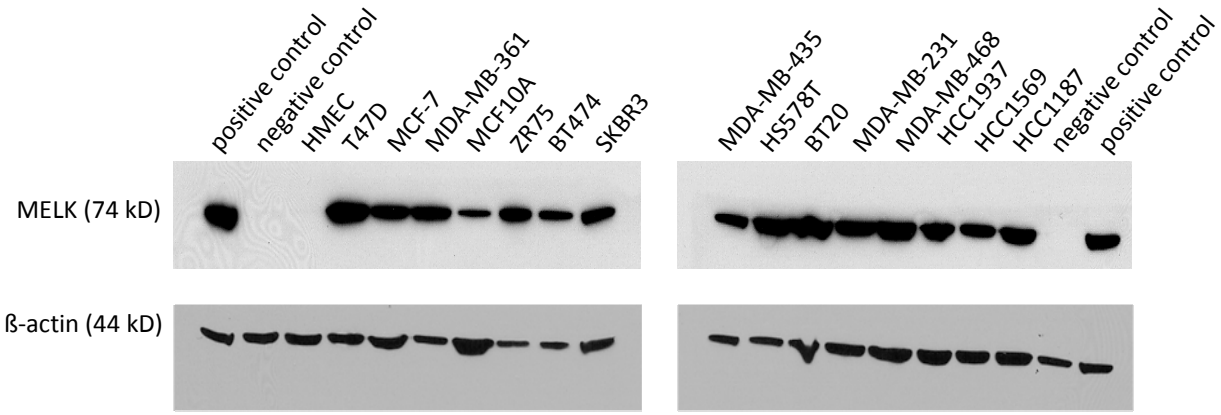
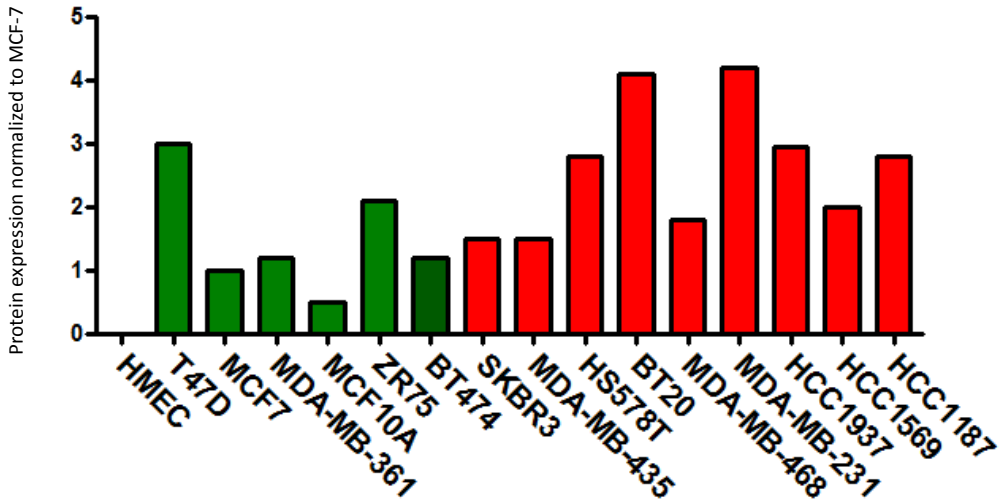
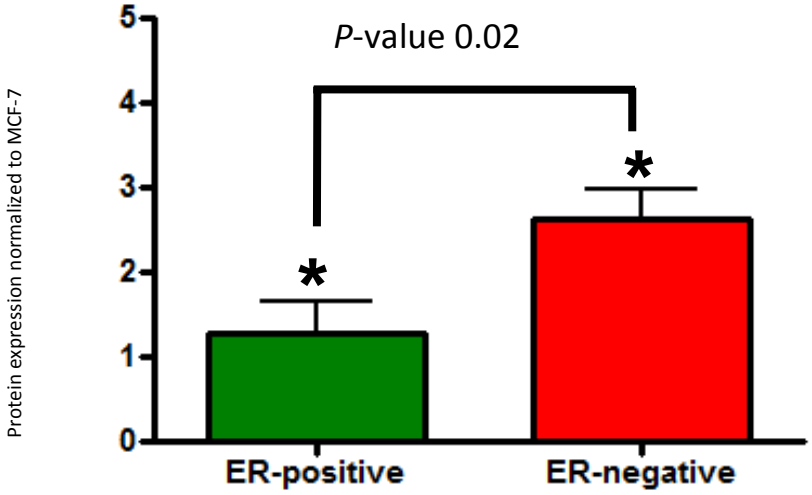

relative fold change normalized to cyclophilin
0
1
2
3
4
5
HCC1569
HCC1954
HCC1937
HS578T
HCC1187
MDA MB-231
BT20
BT549
SKBR3
MDA MB-468
BT474
MCF7
ZR75-1
T47D
Cell lines
0
1
2
3
4
5
6
ER-negative
ER-positive
P-value: 0.007

Figure 5.5 – ER-negative breast cancer cell lines express higher MELK protein levels than ER-positive breast cancer cell lines. MELK protein expression was measured using western blotting and protein levels were normalized using beta actin, measuring pixel intensity in MELK bands normalized to pixel intensity in loading control beta actin bands. Comparison of the two groups using Student's t-test shows that protein expression levels were significantly higher (*P*-value 0.02) in the ER-negative breast cancer cell lines. hTERT transformed HMEC cells overexpressing MELK were used as positive controls and MCF-7 cells with MELK knocked down with siRNA were included as negative controls. Green color indicates ER-positive breast cancer cell lines, red indicates ER-negative breast cancer cell lines. Experiment and analysis done by Corey Speers.

Figure 5. 5


positive control
negative control
HMEC
T47D
MCF-7
MDA-MB-361
MCF10A
ZR75
BT474
SKBR3
MDA-MB-435
HS578T
BT20
MDA-MB-231
MDA-MB-468
HCC1937
HCC1569
HCC1187
negative control
positive control
MELK (74 kD)
ß-actin (44 kD)


Protein expression normalized to MCF-7
0
1
2
3
4
5
HMEC
T47D
MCF7
MDA-MB-361
MCF10A
ZR75
BT474
SKBR3
MDA-MB-435
HS578T
BT20
MDA-MB-468
MDA-MB-231
HCC1937
HCC1569
HCC1187


P-value 0.02
Protein expression normalized to MCF-7
0
1
2
3
4
5
*
*
ER-positive
ER-negative

2 days after transfection in all the cell lines used, and MELK expression remained low until at least day 5 (**Figure 5.6**).

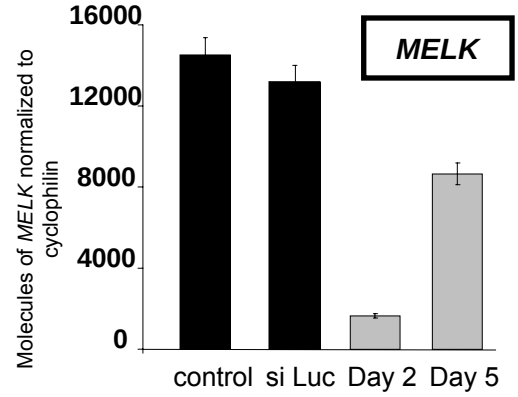
MELK knockdown had significant growth-inhibitory effects in 6 of the 7 ER-negative breast cancer lines (MDA-MB-468, MDA-MB-231, HCC 1937, HCC 1569, HCC 1187, and BT549) but had little or no effect on the ER-positive breast cancer cell lines (slight effect in T47D cells). Representative growth curves are shown in **Figure 5.7.** These experiments show that HCC 1187, HCC 1937, and MDA-MB-468 were most sensitive to MELK inhibition. These are also the cell lines with the highest MELK expression as measured by Q-RT-PCR analysis. These results indicate that MELK expression is indeed critical for the growth of many ER-negative breast cancer cell lines, while ER-positive breast cancer cells have mitogenic signaling pathways that can compensate for the loss of MELK expression.

5.3.8 MELK Expression Predicts Poor Metastasis-Free Survival

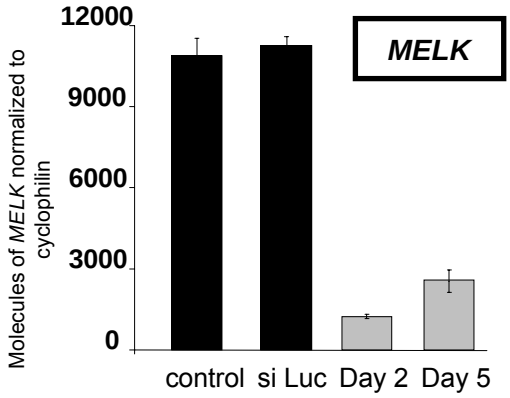
While our studies identified MELK as being significantly overexpressed in ER-negative breast cancers and knockdown of MELK led to significant growth inhibition in ER-negative cell lines, we wanted to determine whether MELK expression was correlated with prognosis in patients with breast cancer. For the prognostic studies we analyzed the survival data from five very different datasets that had substantial clinical follow up data (Wang, van de Vijver, Ivshina, Denmark, and Desmedt,) [10, 11, 17, 22]. While the datasets have been summarized previously in Chapter 2 (**Chapter 2.2.7**), a

Figure 5.6 – siRNA designed against MELK effectively knocks down the expression of MELK at day 2 and day 5. siRNA duplexes designed against MELK shows that transient transfection can decrease expression of MELK RNA by >70% in multiple cell lines. All cell lines used in the siRNA growth studies showed at least 70% inhibition at day 2. Data are represented as mean ± SD. Experiment and analysis done by Corey Speers.

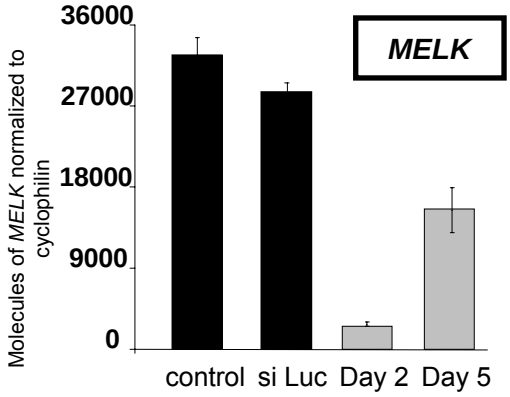
Figure 5.6



MDA-MB-468 cells



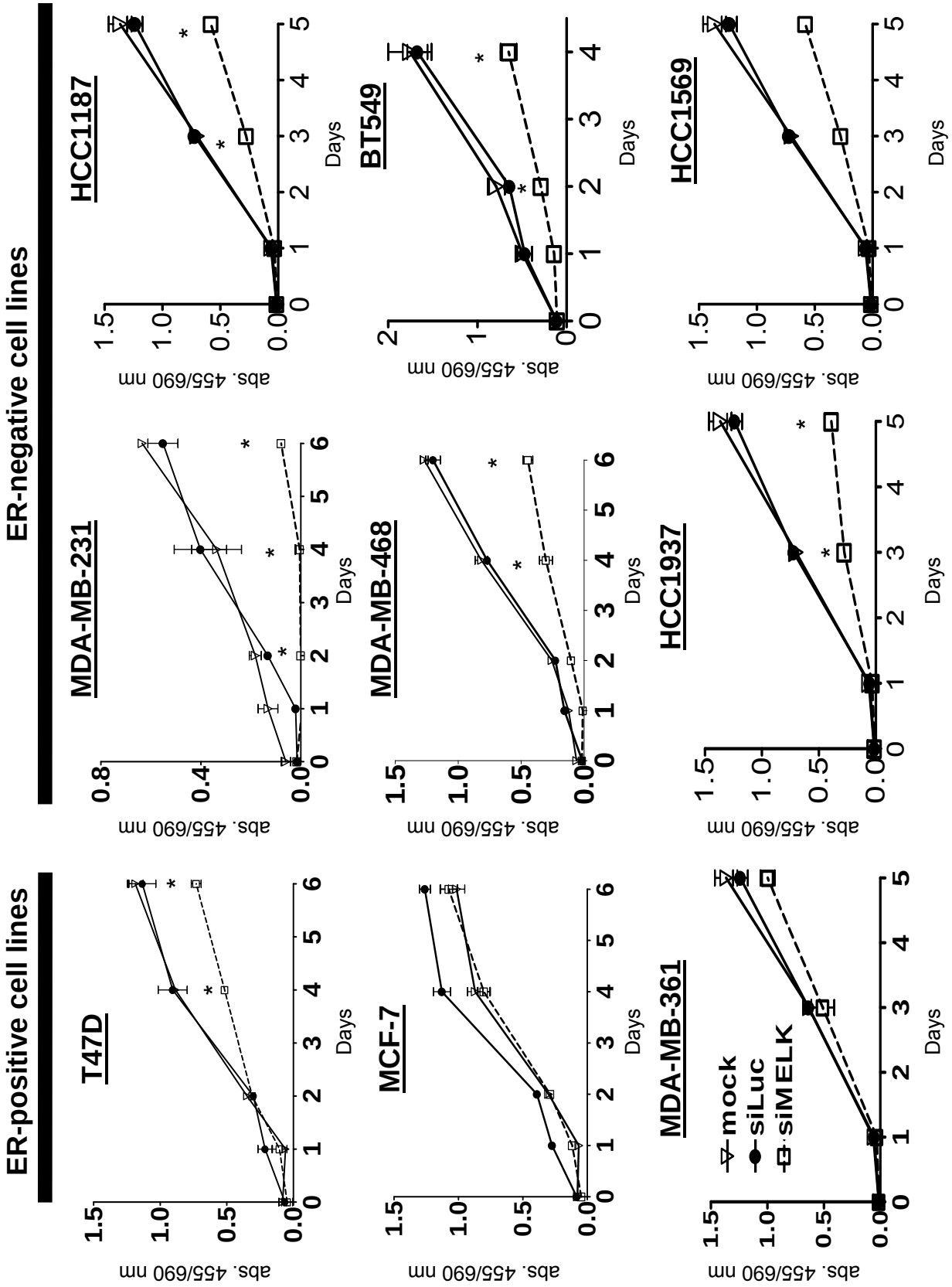
MDA-MB-231 cells



HCC1187 cells

Figure 5.7- Effect of MELK siRNA knockdown on the growth of ER-negative and ER-positive breast cancer cells. MELK knockdown inhibited growth in the ER-negative breast cancer cell lines (HCC 1937, HCC 1569, HCC 1187, BT549, MDA-MB-468 and MDA-MB-231) but not in the ER-positive breast cancer cell lines (MDA-MB-361, and MCF-7). ER-positive T47D cells were slightly inhibitied by MELK knockdown. Asterisk denotes significant difference in curves between kinase of interest knockdown and mock transfected growth curves, *P*-value < 0.05. Data are represented as mean ± SD. Experiment and analysis done by Corey Speers.

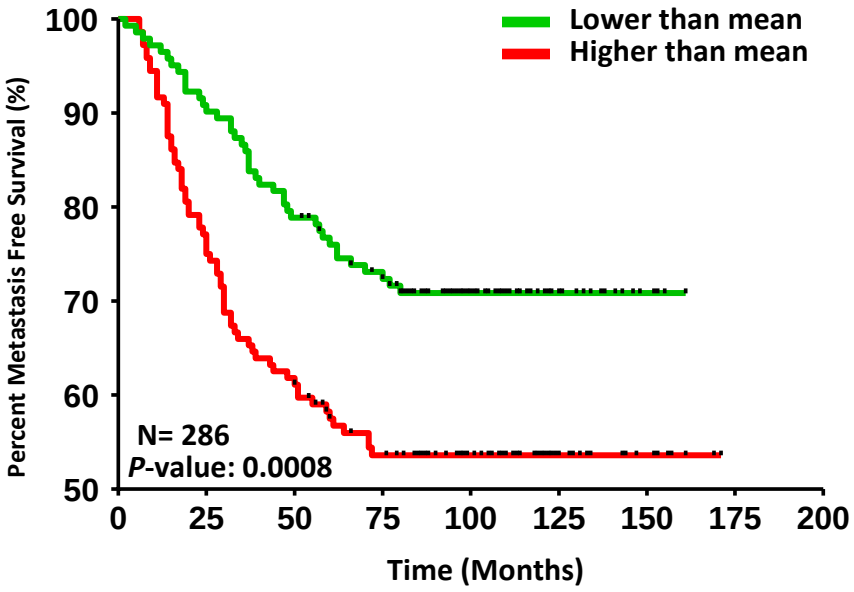
Figure 5.7

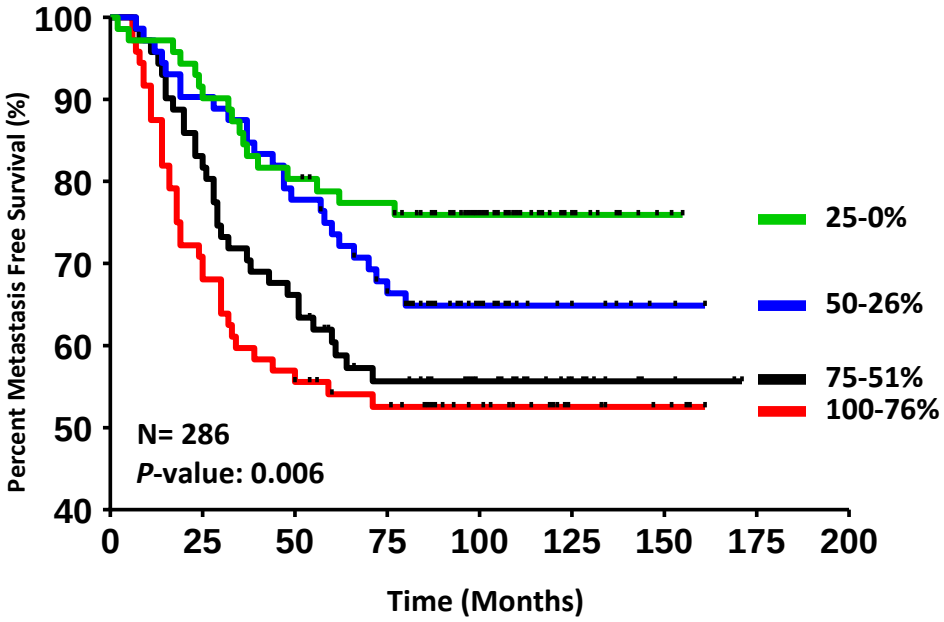

ER-positive cell lines
ER-negative cell lines
T47D
MCF-7
MDA-MB-361
MDA-MB-231
MDA-MB-468
HCC1187
BT549
HCC1937
HCC1569
mock
siLuc
siMELK
abs. 455/690 nm
Days

brief description follows. The Wang dataset was obtained using breast cancer samples from patients with lymph-node negative breast cancer who were treated with breast conserving surgery or modified radical mastectomies from 1980–95. These patients also received radiotherapy when indicated, but did not receive systemic chemotherapy or hormonal adjuvant therapy. This time period was also prior to the development of the anti-HER2 therapy, trastuzumab (Herceptin), and these patients were not treated with trastuzumab (Herceptin). 219 patients had breast-conserving surgery and 67 had modified radical mastectomies. Radiotherapy was given to 248 patients (87%), and metastasis free survival was tracked in all patients. In this data set, we first divided tumors by the mean level of MELK expression. Kaplan-Meier analysis of the metastasis-free survival between the different groups (higher than mean versus lower than mean expression) showed that women who had higher than mean expression of MELK had a much worse prognosis than those with lower MELK expression (**Figure 5.8**). In addition, when expression was ordered in descending order and patients were divided into quartiles based on the level of MELK expression, a step-like pattern was discovered in Kaplan-Meier analysis. Those patients whose tumors were in the highest quartile of MELK expression had the worst metastasis-free survival, followed by the second quartile of patients (75-50%), who had the second worst outcomes. Women in the third quartile had better survival, while women in the lowest quartile whose tumors had the lowest expression of MELK had the best survival (**Figure 5.8**). To investigate whether MELK expression was correlated with overall survival, we examined the van de Vijver dataset

Figure 5.8 – Kaplan-Meier analysis in the Wang dataset shows women whose tumors have higher than mean expression of MELK have significantly worse metastasis-free survival. Higher MELK expression is correlated with significantly worse metastasis-free survival, both when expression is divided at the mean or into quartiles. Overall *P*-value comparing the curves is listed. Analysis done by Corey Speers with input with Dr. Susan Hilsenbeck.

Figure 5.8. MELK Expression and Metastasis-Free Survival: Wang dataset



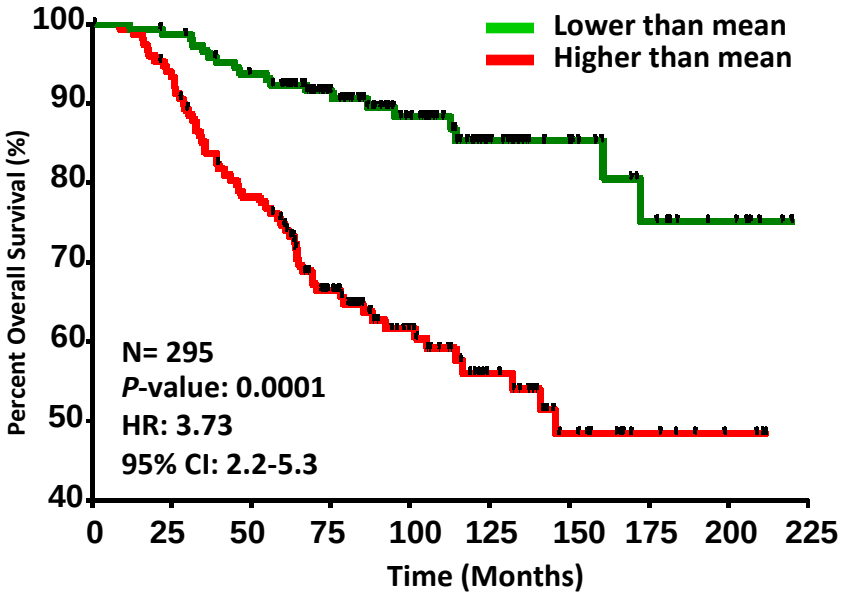


[10] and found similar results. In this dataset, all patients had stage I or II breast cancer and were younger than 53 years old; 151 had lymph-node–negative disease, and 144 had lymph node–positive disease. Ten of the 151 patients who had lymph-node–negative disease and 120 of the 144 who had lymph-node–positive disease received adjuvant systemic therapy consisting of chemotherapy (90 patients), hormonal therapy (20), or both (20). As seen with the Wang dataset, patients whose tumors had high expression of MELK had a significantly worse overall survival than those with low expression of MELK (**Figure 5.9**). The same discriminatory step-wise pattern was noted when patients were divided into quartiles based on the expression of MELK in their tumors (**Figure 5.9**). Hazards ratios and 95% confidence intervals were also calculated for all of the analyses and are listed.

The Ivshina dataset contains data from 307 tumors from a study initiated to evaluate the differences in gene expression profiles between good and poor outcome grade 2 breast tumors (see **Supplementary Table 5.1.1** for complete dataset characteristics). This relatively large dataset had clinical follow-up data that allowed for the analysis of MELK expression and its correlation with clinical outcome. As with the other datasets, patients whose tumors had high expression of MELK had a significantly worse overall survival than those with low expression of MELK (**Figure 5.10**). The same discriminatory step-wise pattern was noted when patients were divided into quartiles based on the expression of MELK in their tumors (**Figure 5.10**). This association of high MELK expression in patients' tumors and poor overall survival was validated in two

Figure 5.9 – Kaplan-Meier analysis in the vande Vijver dataset shows women whose tumors have higher than mean expression of MELK have significantly worse overall survival. Higher MELK expression is correlated with significantly worse overall survival, both when expression is divided at the mean or into quartiles. Overall *P*-value comparing the curves is listed. HR refers to the hazards ratios and 95% CI refers to the confidence intervals. Analysis done by Corey Speers with input with Dr. Susan Hilsenbeck.

Figure 5.9. MELK Expression and Overall Survival: vande Vijver dataset



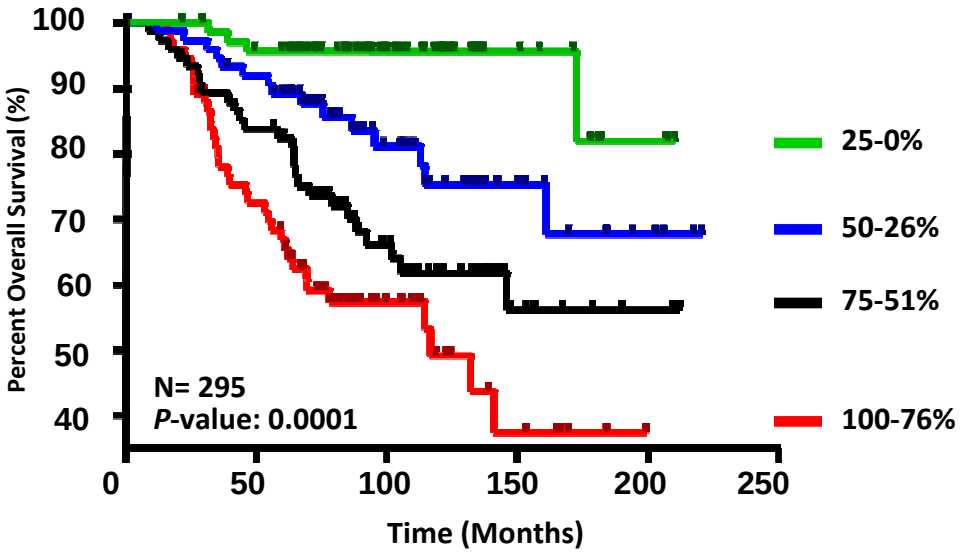
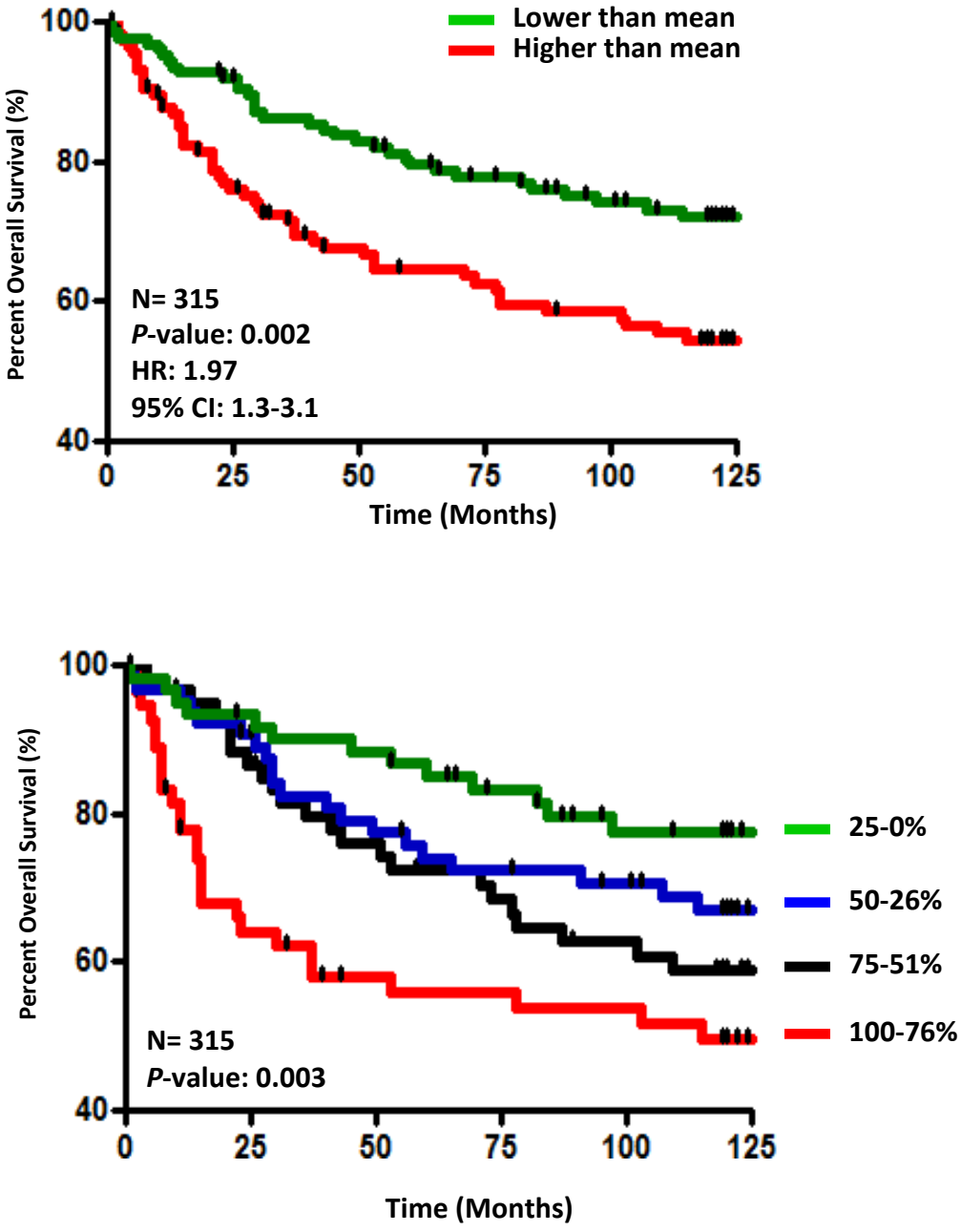


Figure 5.10 – Kaplan-Meier analysis in the Ivshina dataset shows women whose tumors have higher than mean expression of MELK have significantly worse overall survival. Higher MELK expression is correlated with significantly worse overall survival, both when expression is divided at the mean or into quartiles. Overall *P*-value comparing the curves is listed and is based on the Chi squared score. HR refers to the hazards ratios and 95% CI refers to the confidence intervals. Analysis done by Corey Speers with input with Dr. Susan Hilsenbeck.

Figure 5.10. MELK Expression and Overall Survival: Ivshina dataset



additional datasets (Denmark and Desmedt, **Figure 5.11** and **Figure 5.12**, respectively; for complete dataset characteristics see **Supplementary Table 5.1.2**). These data suggest that MELK expression may be prognostic and identify patients who will have better or worse outcomes based on the level of MELK expression, and may be clinically useful in identifying patients requiring more aggressive clinical management.

5.3.9 MELK Expression is Prognostic in Breast Cancer

Though Kaplan-Meier analysis in multiple datasets suggested that MELK expression may in itself be prognostic, we performed uni- and multivariate Cox regression analysis to determine if MELK expression was independently prognostic, including all available biological characteristics of the tumor in the model (excluding treatment modalities). For this analysis, we used the training set of tumors from Denmark that has previously been described (see **Chapter 4**). This dataset was chosen because of its relatively large size, high percentage of aggressive tumors, and extensive clinical data. MELK expression was measured by gene expression profiling and normalized using a QC-RMA method. MELK was analyzed as a continuous variable, and expression values ranged from 7 to 15, with mean expression of 11.91 in all tumors. In a multivariate Cox proportional hazards model, high MELK expression was shown to be independently associated with a poor prognosis, (HR=1.19 95% CI: (1.05, 1.36) p=0.007). Thus, each unit increase in MELK expression resulted in a 19% increased risk of death, that is patients whose tumors had a MELK expression value of 13 had an 76%

Figure 5.11 – Kaplan-Meier analysis in the Denmark dataset shows women whose tumors have higher than mean expression of MELK have significantly worse overall survival. Higher MELK expression is correlated with significantly worse overall survival, both when expression is divided at the mean or into quartiles. Overall *P*-value comparing the curves is listed and is based on the Chi squared score. HR refers to the hazards ratios and 95% CI refers to the confidence intervals. Analysis done by Corey Speers with input with Dr. Susan Hilsenbeck.

Figure 5.11. MELK Expression and Overall Survival: Denmark dataset

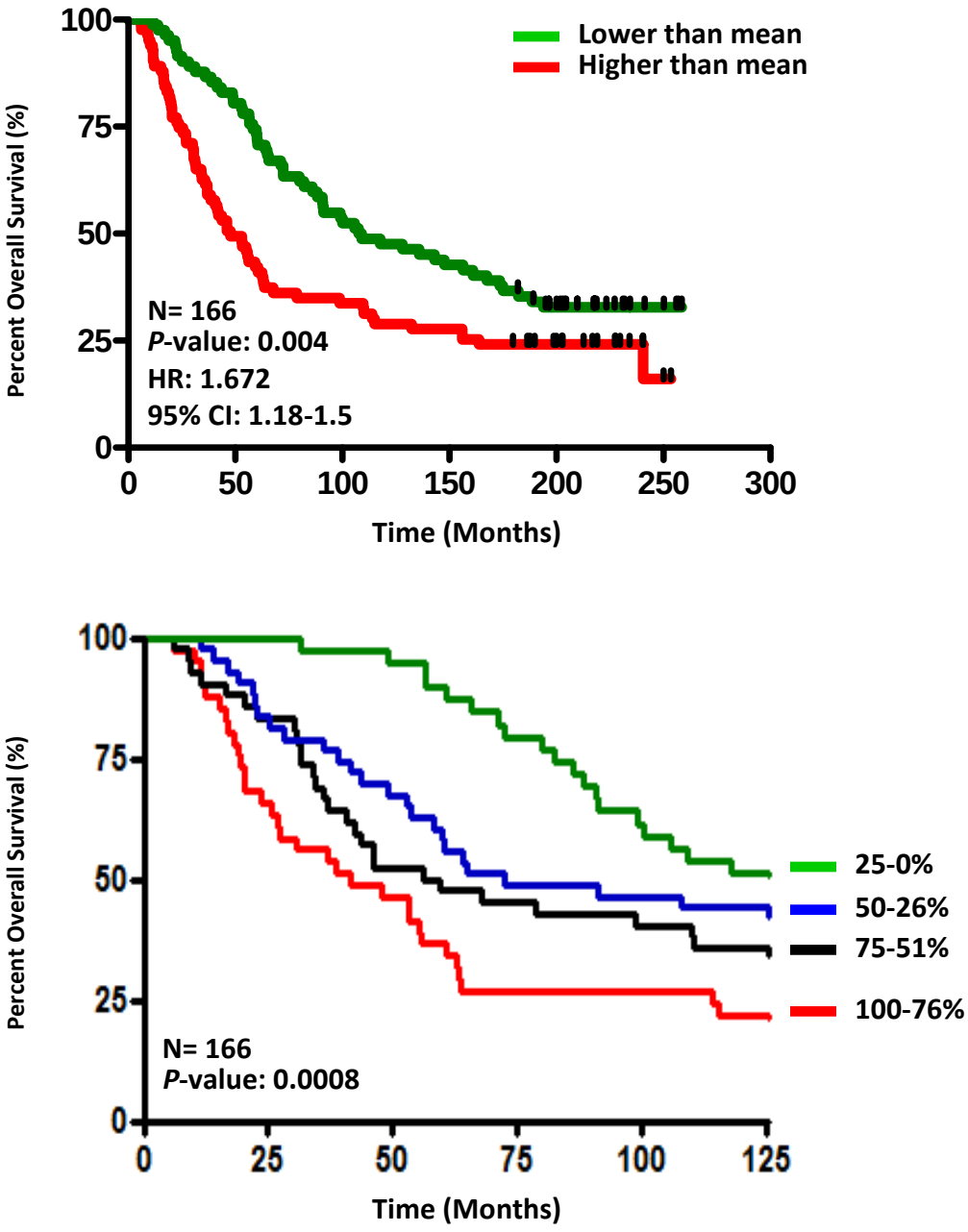
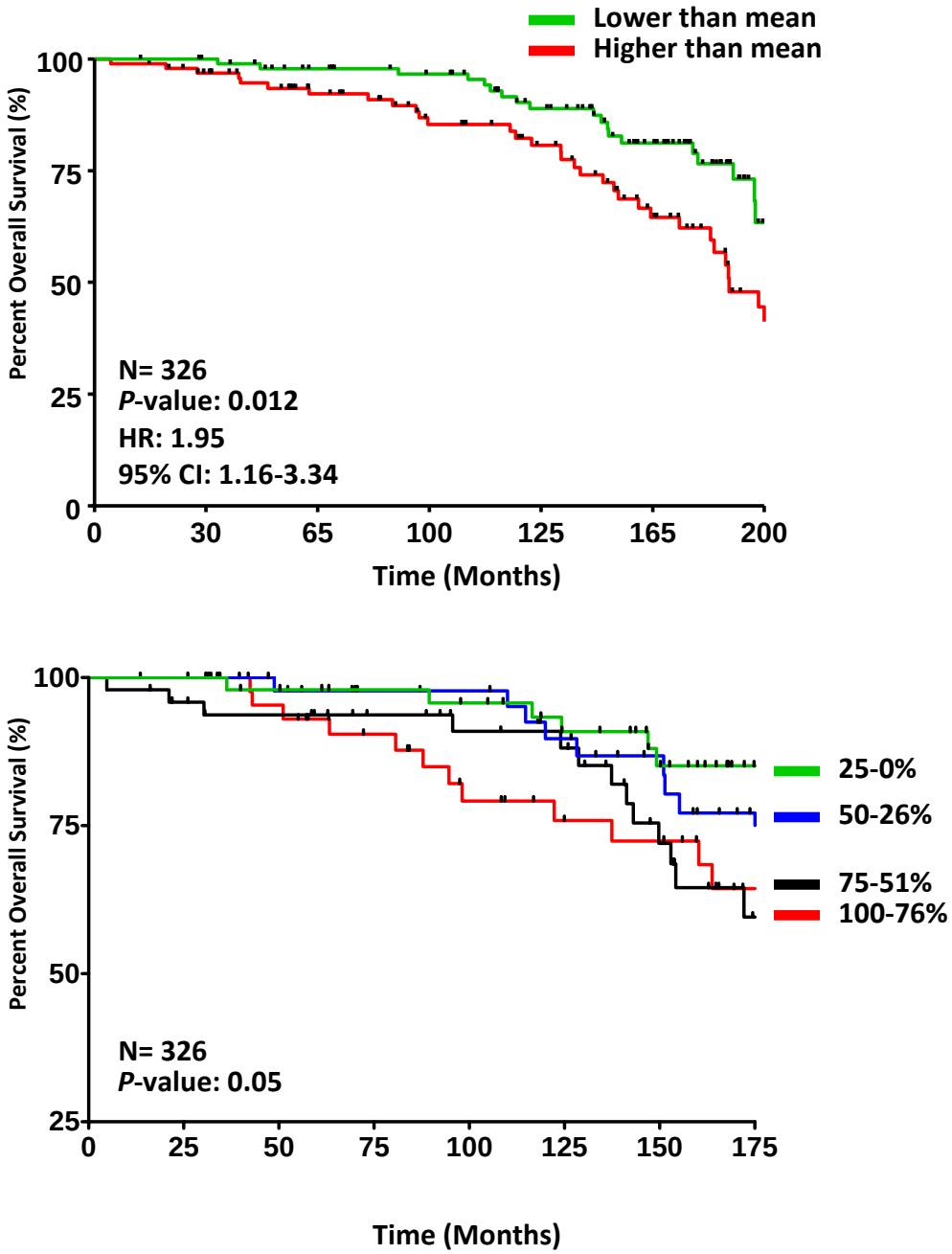


Figure 5.12 – Kaplan-Meier analysis in the Desmedt dataset shows women whose tumors have higher than mean expression of MELK have significantly worse overall survival. Higher MELK expression is correlated with significantly worse overall survival, both when expression is divided at the mean or into quartiles. Overall *P*-value comparing the curves is listed and is based on the Chi squared score. HR refers to the hazards ratios and 95% CI refers to the confidence intervals. Analysis done by Corey Speers with input with Dr. Susan Hilsenbeck.

Figure 5.12. MELK Expression and Overall Survival: Desmedt



increased risk of death compared to patients whose tumors had a MELK expression value of 9. ER status also significantly influenced survival. Patients with ER-positive tumors were more likely to survive than patients with ER-negative tumors (HR=0.62, 95% CI: 0.39-0.99, p=0.045). Though nodal status, which has previously been shown to be independently prognostic in breast cancer, did not reach statistical significance in this model, this dataset included patients that were 93% nodal positive, meaning that almost all the patients were LN-positive and thus variance was insufficient for statistical significance. These data indicated that MELK expression, in addition to ER status, are independently prognostic in this dataset and that high expression of MELK is associated with an increased risk of death. The multivariate analysis and modeling was done by Krystal Sexton at Baylor College of Medicine.

5.4 Discussion

In this report we show that gene expression profiling of human breast tumors is able to identify MELK as a kinase more highly expressed in ER-negative breast cancers as compared to ER-positive breast cancers. Further analysis revealed that MELK is not normally expressed at appreciable levels in most normal tissues, but is highly expressed in cancerous tissues, including breast cancer tissue. Analysis of publicly available breast tumor data sets confirmed that MELK is indeed highly expressed in ER-negative breast cancer. Additional validation in breast tumor samples as well as breast cancer cell lines confirms MELK overexpression at the RNA and protein levels. Furthermore, studies in which knockdown of MELK using siRNA demonstrated that it is critical for the growth of 7 of 8 ER-negative breast cancer cell lines tested. Analysis of MELK expression in human breast tumors demonstrated that patients whose tumors have high expression of this kinase have significantly poorer outcomes than patients with low expression of MELK. Finally, multivariate analysis demonstrated that MELK is an independent prognostic factor associated with poor prognosis. Such results suggest that women whose tumors have high MELK expression have a poor prognosis and may benefit from aggressive treatment. In addition, this study identifies MELK itself as a potential target for the treatment of ER-negative breast cancer.

Currently effective target therapies for ER-positive breast cancer exist. These therapies include anti-estrogens (anti ERα drugs) such as tamoxifen and fulvestrant, and more recently aromatase inhibitors, that have led to significant improvements in the

overall survival of patients with ER-positive breast cancer [23-27]. Even for ER-positive and ER-negative breast tumors that overexpress HER2/*neu* (20-30% of breast cancers), trastuzumab and more recently lapatinib drugs, which both interfere with HER2 signaling, have resulted in a marked improvement in both response and survival [28-34]. Unfortunately, only 30-60% of women with HER2-positive tumors benefit from trastuzumab [32-34], and these therapies are of no use in patients whose tumors don't overexpress HER2. The development of effective targeted therapies in ER-negative breast cancer has been hindered by a lack of verified targets.

Recent efforts have led to the identification of some potential targets in ER-negative breast cancer. Peptide growth factors or their receptors are currently being investigated as possible targets for the treatment of this kind of breast cancer. These include the epidermal growth factor receptor (EGFR), the insulin-like growth factor receptor (IGFR), fibroblast growth factor receptors (FGFR), vascular endothelial growth factor (VEGF). In preclinical studies these pathways have been shown to be active in breast cancer cells, and inhibitors of these pathways are being tested in both preclinical and clinical trials [25, 28, 30, 31, 35-37]. Additional targeted therapies, including PI3-kinase inhibitors, CHEK1 inhibitors, and Src inhibitors, are in various stages of clinical trials in breast cancer [4, 38, 39]. Though these therapies may prove effective in treating subsets of women with breast cancer, it is clear that additional therapies are critically needed. Many of these therapies have side effects that limit their clinical utility, and the problem of drug resistance remains a substantial limitation to their use.

Additionally, though these targets hold promise for the treatment of tumors that express the aforementioned markers, many ER-negative tumors do not express any of these targets. Thus, it is critical that additional, novel targets for the treatment of ER-negative breast cancer be developed.

MELK may be one such target. MELK is an atypical member of the snf1/AMPK family of serine/threonine kinases [40]. This family is largely associated with cell survival under conditions of environmental challenge, such as nutrient starvation [41, 42]. Previous studies, however, have demonstrated that MELK may regulate other important processes, like stem cell self renewal through control of the cell cycle [43]. Likewise, MELK has been identified as a cell cycle modulator in tumor cell lines and was recently identified as an important target for certain somatic tumors, including colorectal, lung, and ovarian cancers [44]. Here we report that MELK is not expressed in most normal tissues, is more highly expressed in ER-negative breast cancers compared to ER-positive breast cancers, and is critical for ER-negative breast cancer cell line growth. MELK is also independently prognostic in human breast cancer, and high MELK expression is significantly associated with poor overall survival. The only other genes shown to be independently prognostic in breast cancer (ER and HER2) are currently targeted for the effective treatment of breast cancer. Thus, the results reported herein demonstrate that MELK may be an ideal target for the effective treatment of aggressive ER-negative breast cancers. Given the current difficulty in treating ER-negative breast cancer, the identification of this kinase, shown to be critical for the growth of these

cancers and independently prognostic, represents the first step towards developing and utilizing additional targets for the treatment of these poor prognosis ER-negative breast cancers.

References

1. Carey, L.A., et al., *Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.* Jama, 2006. **295**(21): p. 2492-502.
2. Swaby, R.F., C.G. Sharma, and V.C. Jordan, *SERMs for the treatment and prevention of breast cancer.* Rev Endocr Metab Disord, 2007. **8**(3): p. 229-39.
3. Tan, A.R. and S.M. Swain, *Adjuvant chemotherapy for breast cancer: an update.* Semin Oncol, 2001. **28**(4): p. 359-76.
4. Miller, K., et al., *Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer.* N Engl J Med, 2007. **357**(26): p. 2666-76.
5. Viani, G.A., et al., *Adjuvant trastuzumab in the treatment of her-2-positive early breast cancer: a meta-analysis of published randomized trials.* BMC Cancer, 2007. **7**: p. 153.
6. Dowell, J.E. and J.D. Minna, *EGFR mutations and molecularly targeted therapy: a new era in the treatment of lung cancer.* Nat Clin Pract Oncol, 2006. **3**(4): p. 170-1.
7. Herbst, R.S., M. Fukuoka, and J. Baselga, *Gefitinib--a novel targeted approach to treating cancer.* Nat Rev Cancer, 2004. **4**(12): p. 956-65.
8. Minna, J.D. and J. Dowell, *Erlotinib hydrochloride.* Nat Rev Drug Discov, 2005. **Suppl**: p. S14-5.
9. Dancey, J. and E.A. Sausville, *Issues and progress with protein kinase inhibitors for cancer treatment.* Nat Rev Drug Discov, 2003. **2**(4): p. 296-313.
10. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.
11. Wang, Y., et al., *Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.* Lancet, 2005. **365**(9460): p. 671-9.
12. Miller, L.D., et al., *An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival.* Proc Natl Acad Sci U S A, 2005. **102**(38): p. 13550-5.
13. Hess, K.R., et al., *Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer.* J Clin Oncol, 2006. **24**(26): p. 4236-44.
14. Yu, K., et al., *A modular analysis of breast cancer reveals a novel low-grade molecular signature in estrogen receptor-positive tumors.* Clin Cancer Res, 2006. **12**(11 Pt 1): p. 3288-96.
15. Richardson, A.L., et al., *X chromosomal abnormalities in basal-like human breast cancer.* Cancer Cell, 2006. **9**(2): p. 121-32.
16. Chin, K., et al., *Genomic and transcriptional aberrations linked to breast cancer pathophysiologies.* Cancer Cell, 2006. **10**(6): p. 529-41.
17. Ivshina, A.V., et al., *Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer.* Cancer Res, 2006. **66**(21): p. 10292-301.

18. Ginestier, C., et al., *Prognosis and gene expression profiling of 20q13-amplified breast cancers.* Clin Cancer Res, 2006. **12**(15): p. 4533-44.
19. Minn, A.J., et al., *Genes that mediate breast cancer metastasis to lung.* Nature, 2005. **436**(7050): p. 518-24.
20. Bild, A.H., et al., *Oncogenic pathway signatures in human cancers as a guide to targeted therapies.* Nature, 2006. **439**(7074): p. 353-7.
21. Whitlock, M.C., *Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach.* J Evol Biol, 2005. **18**(5): p. 1368-73.
22. Desmedt, C., et al., *Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series.* Clin Cancer Res, 2007. **13**(11): p. 3207-14.
23. Patel, R.R., C.G. Sharma, and V.C. Jordan, *Optimizing the antihormonal treatment and prevention of breast cancer.* Breast Cancer, 2007. **14**(2): p. 113-22.
24. Howell, A., et al., *Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer.* Lancet, 2005. **365**(9453): p. 60-2.
25. Thurlimann, B., et al., *A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer.* N Engl J Med, 2005. **353**(26): p. 2747-57.
26. Fisher, B., et al., *Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study.* J Natl Cancer Inst, 2005. **97**(22): p. 1652-62.
27. Cummings, S.R., et al., *The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation.* Jama, 1999. **281**(23): p. 2189-97.
28. Geyer, C.E., et al., *Lapatinib plus capecitabine for HER2-positive advanced breast cancer.* N Engl J Med, 2006. **355**(26): p. 2733-43.
29. Moy, B., et al., *Lapatinib.* Nat Rev Drug Discov, 2007. **6**(6): p. 431-2.
30. Mayer, E.L., N.U. Lin, and H.J. Burstein, *Novel approaches to advanced breast cancer: bevacizumab and lapatinib.* J Natl Compr Canc Netw, 2007. **5**(3): p. 314-23.
31. Sonpavde, G., *Lapatinib plus capecitabine in breast cancer.* N Engl J Med, 2007. **356**(14): p. 1471; author reply 1471-2.
32. Dawood, S., et al., *Efficacy and safety of neoadjuvant trastuzumab combined with paclitaxel and epirubicin: a retrospective review of the M. D. Anderson experience.* Cancer, 2007. **110**(6): p. 1195-200.
33. Burstein, H.J., et al., *Trastuzumab plus vinorelbine or taxane chemotherapy for HER2-overexpressing metastatic breast cancer: the trastuzumab and vinorelbine or taxane study.* Cancer, 2007. **110**(5): p. 965-72.

34. Garrison, L.P., Jr., et al., *Cost-effectiveness analysis of trastuzumab in the adjuvant setting for treatment of HER2-positive breast cancer.* Cancer, 2007. **110**(3): p. 489-98.
35. Cobleigh, M.A., et al., *A phase I/II dose-escalation trial of bevacizumab in previously treated metastatic breast cancer.* Semin Oncol, 2003. **30**(5 Suppl 16): p. 117-24.
36. Dickler, M.N., et al., *A phase II trial of erlotinib in combination with bevacizumab in patients with metastatic breast cancer.* Clin Cancer Res, 2008. **14**(23): p. 7878-83.
37. Miller, K.D., et al., *Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer.* J Clin Oncol, 2005. **23**(4): p. 792-9.
38. Garber, K., *New checkpoint blockers begin human trials.* J Natl Cancer Inst, 2005. **97**(14): p. 1026-8.
39. deGraffenried, L.A., et al., *Inhibition of mTOR activity restores tamoxifen response in breast cancer cells with aberrant Akt Activity.* Clin Cancer Res, 2004. **10**(23): p. 8059-67.
40. Lizcano, J.M., et al., *LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1.* EMBO J, 2004. **23**(4): p. 833-43.
41. Suzuki, A., et al., *ARK5 suppresses the cell death induced by nutrient starvation and death receptors via inhibition of caspase 8 activation, but not by chemotherapeutic agents or UV irradiation.* Oncogene, 2003. **22**(40): p. 6177-82.
42. Kato, K., et al., *Critical roles of AMP-activated protein kinase in constitutive tolerance of cancer cells to nutrient deprivation and tumor formation.* Oncogene, 2002. **21**(39): p. 6082-90.
43. Nakano, I., et al., *Maternal embryonic leucine zipper kinase (MELK) regulates multipotent neural progenitor proliferation.* J Cell Biol, 2005. **170**(3): p. 413-27.
44. Gray, D., et al., *Maternal embryonic leucine zipper kinase/murine protein serine-threonine kinase 38 is a promising therapeutic target for multiple cancers.* Cancer Res, 2005. **65**(21): p. 9751-61.

Supplementary Table 5.1.1 – Ivshina dataset					
	Uppsala n=249			**Stockholm n=58**	**Singapore n=40**
Variables, by grade	**G1** n=68	**G2** n=126	**G3** n=55	**G2** n=58	**G2** n=40
Age, median yrs	62	63	62	58	52
<55 years, %	26	25	44	41	60
Tumor size, cm	1.8	2.2	2.9	2.5	2.8
Nodes, positive, %	15	35	55	50	40
ER negative tumors, %	3	9	38	7	28
Follow up, median yrs	11	9	6	7	-
All recurrences, %	26	39	50	24	-
Endocrine therapy, %	18	37	36	62	-
Chemotherapy, %	4	6	22	5	-
Combine therapy, %	2	3	0	16	-
No systemic therapy, %	77	54	45.5	17	

Supplementary Table 5.1.1 – Clinical characteristics of the tumors in the Ivshina dataset. This study was designed to look at gene expression differences between good and poor outcome Grade 2 tumors. For the purposes of analyzing the effect of MELK expression on outcome, only tumors from the Uppsala and Stockcolm cohorts were used as they were the tumors with clinical follow up data available. Analysis was done by Corey Speers with input from Dr. Susan Hilsenbeck.

Supplementary Table 5.1.2 – Denmark dataset

Characteristic		Training Set N=166 (%)
Age	Mean	54.8 (SD 9.14)
	Range	30-69
Tumor Stage	T1	49 (30%)
	T2	97 (58%)
	T3	20 (12%)
Nodal Status	0 nodes positive	11 (7%)
	1-3 nodes positive	77 (46%)
	>3 nodes positive	78 (47%)
Recurrence Status	No local recurrence	124 (75%)
	Local recurrence	42 (25%)
	Unknown	0 (0%)
Metastasis	No distant metastasis	67 (40%)
	Distant metastasis	99 (60%)
Survival Months- All	Mean	107.5 (SD 80.65)
	Range	6-258
ER	Positive	126 (76%)
	Negative	40 (24%)
HER2/Neu	Positive	34 (21%)
	Negative	102 (61%)

Supplementary Table 5.1.2 – Clinical characteristics of the tumors in the Denmark dataset. This study from a Denmark cohort of patients was provided to Dr. Gordon Mills at M.D. Anderson. This set had clinical follow-up data that allowed for the analysis of the effect of MELK expression on outcome. Data generously provided by Dr. Gordon Mills for analysis by Corey Speers.

Chapter 6

Discussion and Future Directions

6.1 Overview

Approximately 30% of breast tumors are estrogen receptor alpha (ER)-negative, but these aggressive breast cancers account for a much greater proportion of deaths than their prevalence would suggest. This is primarily because of poor tumor response to traditional therapies [1]. Breast cancer development and progression involves the interplay between two major classes of growth-promoting agents, steroid hormones (estrogens and progestins) and polypeptide growth factors. These growth factors and their receptors have served as important targets for the treatment of ER-positive and HER2/*neu* (ErbB2) positive breast cancer. The development of anti-estrogens such as tamoxifen and fulvestrant, and more recently aromatase inhibitors, has led to significant improvements in the overall survival of patients with ER-positive breast cancer [2]. Similarly targeting HER2/*neu* (ErbB2), which is overexpressed on 20-30% of breast cancers, with trastuzumab and more recently lapatinib has resulted in a marked improvement in both response and survival [3]. Unfortunately, only 30-60% of women with HER2-positive tumors benefit from trastuzumab, and none of these treatments are useful in treating women with ER-negative, HER2-negative breast cancer.

Peptide growth factors or their receptors, in addition to HER2, are currently being investigated as possible targets for the treatment of breast cancer. These include the epidermal growth factor receptor (EGFR), the insulin-like growth factor receptor (IGFR), fibroblast growth factor receptors (FGFR) and vascular endothelial growth factor (VEGF) [4-6]. In preclinical studies these pathways have been shown to be active in

breast cancer cells, and inhibitors of these pathways are being tested in both preclinical and clinical trials. The FGFR has recently been shown to be a predisposing factor in the development of breast cancer [7]. A blocking antibody to VEGF, bevacizumab, has been shown in combination with paclitaxel to improve progression free survival and showed a trend toward prolonging overall survival for women with metastatic breast cancer [6]. This growth factor inhibitor is now being tested in combination with other chemotherapy agents and in early stage breast cancer. Additionally, EGFR inhibitors such as gefitinib and erlotinib, and the dual EGFR/HER2 inhibitor lapatinib have also shown promise in early clinical trials [8-11], and recently the FDA approved the use of lapatinib for the treatment of HER2-positive metastatic breast cancer.

Despite the potential utility of the above targeted therapies, several factors limit their efficacy. First, the patient profile that would benefit from these drugs has yet to be clearly defined and even those that do initially respond often develop resistance to the therapy. Second, the side effect profile of many of these drugs limit their clinical utility. Though effective, many of these drugs have side effects that are themselves too toxic for use in the clinic. Finally, even if the aforementioned drugs are effective and have acceptable toxicity profiles, it is apparent that they will only be effective in a small percentage of patients. Virtually none of these therapies are hypothesized to have clinical utility in patients with ER-negative, HER2-negative breast cancer, especially those with “triple-negative” (ER-negative, PR-negative, HER2-negative) breast cancer.

Thus, it is critical that additional targets be identified and drugs against those targets be developed for the treatment of ER-negative breast cancer.

The overarching goal of this thesis was to identify and validate therapeutic targets unique to ER-negative breast cancer cells that can be the target of directed therapies. To identify such molecules we used gene expression microarray technologies and novel proteomic approaches to identify molecules that play a role in the development and progression of ER-negative breast cancers. To identify novel targets for the treatment of ER-negative breast cancer we hypothesized that breast tumor genomic and proteomic profiling could be used to reveal molecules for the treatment of ER-negative breast cancer.

The work reported in this thesis describes the results of experiments designed to test this hypothesis. In **chapter 3** we used gene expression microarray profiling to identify a distinct kinase gene expression profile that identifies ER-negative breast tumors and subsets ER-negative breast tumors into 4 distinct subgroups. Based upon the types of kinases expressed in these clusters, we identified a cell cycle regulatory group, a S6 kinase pathway group, an immunomodulatory kinase expressing group, and a MAPK pathway group. Furthermore, we showed that this specific kinase profile is validated using independent sets of human tumors, and is also seen in a panel of breast cancer cell lines. Kinase expression knockdown studies showed that many of these

kinases are essential for the growth of ER-negative, but not ER-positive, breast cancer cell lines. Finally, survival analysis of patients with ER-negative breast cancer showed that patients in the S6 kinase pathway signature subgroup had an extremely poor prognosis, while patients whose tumors express high levels of immunomodulatory kinases had significantly better overall survival. This study identified a list of kinases that were prognostic and may serve as druggable targets for the treatment of ER-negative breast cancer.

In **chapter 4**, we described the use of reverse phase proteomic assay (RPPA) technologies to identify proteins, phosphoproteins, and activated pathways that were elevated in ER-negative breast cancers. Again, we demonstrated that ER-negative tumors could be subdivided into four distinct subgroups (ER-low, stathmin and phospho-RB high, S6 kinase-activated, and HER2-activated) based on the expression of these proteins, and that these different subgroups had distinct prognostic profiles. We also identified protein signatures that were associated with a particularly poor prognosis. Finally, we correlated specific proteomic signatures with previously described breast cancer subtypes identified by transcriptional profiling in human breast cancers and identified proteins associated with these distinct intrinsic subtypes of human breast cancer. These results identified proteins and pathways that are activated in specific subsets of ER-negative breast cancers that now serve as targets of future drug development for effective treatment of ER-negative breast cancer.

In **chapter 5** we investigated the growth regulatory activity of one of the kinases identified in **chapter 3** that was expressed more highly in ER-negative breast cancer. This kinase, maternal embryonic leucine-zipper kinase (MELK), is more highly expressed in ER-negative breast cancers as compared to ER-positive breast cancers. It was also more highly expressed in ER-negative breast cancer cell lines, and MELK expression knockdown studies showed MELK is essential for the growth of most ER-negative, but not ER-positive, breast cancer cell lines. Finally, survival analysis of patients with breast cancer shows that those patients whose tumors have high expression of MELK have a significantly poorer prognosis than patients with low expression of MELK, and that MELK is itself and independent prognostic factor in breast cancer with high expression conferring a high probability of early relapse and early death. This study identifies MELK as a particularly promising target for the directed treatment of these highly aggressive, MELK-positive, ER-negative breast cancers.

6.2 Implications of these Studies

The work reported in this thesis extends our current understanding of ER-negative breast cancer and identifies novel targets for the treatment of this deadly disease. Previously, ER-negative breast cancers were thought of as a homogenous group of tumors, all lacking the expression of the estrogen receptor, but otherwise similar. These studies show at both the RNA and protein levels, that significant heterogeneity exists between these ER-negative tumors. Our studies show, for the first time, that these ER-negative tumors can be subdivided into distinct subgroups based on the level and type of kinases or proteins expressed. Furthermore, our studies show that the patients that can be divided into these different subgroups based on their tumors' gene or protein expression profile have very different outcomes. These results have tremendous implications for the identification of patients with ER-negative breast cancer that have a good prognosis and also for those patients with ER-negative breast cancer that have a particularly poor outcome, who may require particularly aggressive treatment. In addition, I have identified a particular protein kinase target, MELK, that is both a prognostic marker and an important potential therapeutic target in triple-negative breast cancer. This work extends the current understanding in the field and will allow for a more personalized view of ER-negative breast cancer, identifying subgroups with different outcomes so that ER-negative cancer will no longer be discussed in such "negative" terms.

This work also extends our knowledge of the basic expression patterns of molecules in ER-negative breast cancer. ER-negative breast cancer is currently defined in terms of the genes and proteins that it does NOT express. This thesis switches the paradigm and identifies genes and proteins that ARE expressed in these tumors. It is no longer necessary to describe these tumors as lacking the expression of estrogen receptor, progesterone receptor, or HER2. Instead, these tumors can be described as expressing an activated S6 kinase pathway, Src, or MELK. Such markers can subgroup these tumors and are putative targets of treatment.

One of the novel findings is the identification of subtypes within ER-negative breast cancer. Both gene expression profiling and proteomic analyses were able to identify 4 subtypes of ER-negative disease. One of these subtypes of ER-negative breast cancer identified by the gene expression profiling studies was the immunomodulatory subtype of ER-negative breast cancer. The role of the immune system in cancer has historically been viewed rather myopically, with investigation into how the immune system itself responds to the "foreign" cancer as the primary focus. It is now being appreciated that the tumor itself may act autonomously to influence the stromal microenvironment and evade recognition by the immunosurveillance machinery. Recent work by Teschendorff *et al*. has also identified an immunomodulatory profile in ER-negative breast cancer which was shown to confer better prognosis [12]. Other groups have published conflicting results on the impact of immunomodulatory genes and what, if any, role they play in the development and prognosis of breast cancer [13,

14]. Future studies will need to investigate whether modulation of intrinsic gene expression by the tumor is an important mechanism by which cancer cells can avoid immunosurveillance, including the proper controls meant to keep aberrant growth in check [12, 13].

The goal of this thesis was to identify additional novel targets for the treatment of ER-negative breast cancer. The gene expression profiling studies detailed in **chapter 3** highlight some of the kinases that may be attractive targets for the treatment of ER-negative breast cancer. **Chapter 4** extends this work and identifies not only proteins that are more highly expressed in ER-negative disease, but identifies pathways that are also activated in ER-negative breast cancer. It is clear from these studies that proteins involved in cell growth, cell cycle regulation, metastasis and invasion, and apoptosis are elevated in ER-negative tumors. Two cyclins (cyclin E and cyclin A) are especially elevated and represent potential therapeutic targets. PAI1, identified as being more highly expressed in ER-negative disease, represents another such target. Recent work developing PAI1 inhibitors have shown promise, and these studies provide the rationale for using these drugs in models of ER-negative breast cancer [15]. Additionally, the identification of the S6 kinase signaling pathway as being activated in ER-negative breast cancer represents another promising target of directed inhibition in ER-negative breast cancer.

The relevance of maternal embryonic leucine zipper kinase (MELK) expression has previously been unappreciated in breast cancer biology. These studies identify

MELK as a potentially important target for the treatment of this disease. My results suggest that MELK is an important growth regulator in ER-negative breast cancer cells that should be easily targeted with small molecule inhibitors. As very few potential targets for the treatment of ER-negative disease exist, the identification of a subset of ER-negative breast cancers that highly express this critical growth regulatory kinase represents a significant advance in ER-negative breast cancer target discovery.

The work described in this thesis identifies genes, proteins, and pathways which may be particularly attractive targets for the treatment of ER-negative breast cancer. These studies identify novel subsets of ER-negative breast cancer, providing evidence for the previously underappreciated clinical heterogeneity in this group of tumors, and demonstrate that patients whose tumors fall into these subtypes have different prognoses. This work utilizes both transcriptomic and proteomic technologies to advance our understanding of the genes and proteins that are over-represented in ER-negative breast cancer, and here we identify many genes that are critical for the growth of ER-negative, but not ER-positive breast cancer. Finally, these studies demonstrate that at least one of these identified targets, MELK, is independently prognostic and may represent a novel and particularly attractive target for the treatment of ER-negative breast cancer.

6.3 Future Directions and Unanswered Questions

The work described in this thesis answers some questions, but also raises even more. Though I have made significant progress in the identification of targets for the treatment of ER-negative breast cancer, these studies lay a foundation for future drug development. With advances in experimental design, and, more importantly, technology, the genomic era is now ushering in the age of systems biology that will aid biological understanding of cancer and pharmaceutical drug development. Further advances in understanding critical growth regulatory pathways of ER-negative breast cancer will come about as genomic, transcriptomic, and proteomic data is integrated and applied to complex questions involving treatment and resistance.

One future direction of these studies is to obtain DNA sequence information in ER-negative breast cancer and perform complex overlap analysis of the changes in DNA, RNA, and proteins in breast cancer. Indeed, DNA copy number determination using comparative genomic hybridization (CGH) or comparative single nucleotide polymorphism analysis (comparative SNP analysis) has shown that breast cancers harbor many gene deletions or gene amplification and that these regions of DNA copy number alteration identify genes or groups of genes that are involved in the oncogenic process [16-21]. The well known breast cancer oncogenes HER2/*neu* and c-MYC, as well as more recently defined oncogenes Rab25 [22], NRG1 [23, 24], and LSM1 [24], have been identified using these DNA-based techniques. CGH and comparative SNP analysis also can identify regions of DNA loss, typically occurring at the site of important tumor

suppressor genes. These techniques have identified the p53 and PTEN tumor suppressor genes specifically in breast cancer (both previously known tumor suppressor genes) [25, 26], as well as novel breast cancer tumor suppressor genes such as PTK2b [27] and BRIT1 [27], and several other DNA regions in which tumor suppressor genes are thought to be located. Future experimental design will need to incorporate these DNA-based technologies so that changes in DNA copy number and sequence can be added to changes in RNA and proteins to obtain a comprehensive 3-dimensional systems biologic fingerprint of breast cancers.

This gene expression microarray analysis (transcriptional profiling) has been extensively used to subtype cancers, predict prognosis and disease free survival, and determine optimal treatment [28-35]. Several landmark transcriptional profiling studies demonstrated the validity of the technique in building a clinically useful molecular taxonomy of breast cancers with a similar histological appearance. Clinically relevant novel subgroups within the ER-positive and ER-negative breast cancers have also been identified, including in this thesis, reflecting the vastly different biology inherent in these tumor subtypes [29, 32]. Thus, identification of molecular markers from transcriptional profiling holds great promise for refining our ability to accurately diagnose and treat breast cancer. Additionally, groups have already used expression profiling to identify gene signatures of chemotherapeutic resistance [30, 36, 37]. Future work will require the integration of very large and very complex datasets that combine DNA sequence and RNA expression data into a useful snapshot of the genomic aberrations in ER-

negative breast cancer. Such information will certainly implicate certain chromosomal regions and genes that are responsible for carcinogenesis, drug resistance, metastasis, and ultimately lead to targetable regions for the treatment of ER-negative breast cancer.

DNA sequence and RNA expression information has been used profile ER-negative breast cancers, but the work reported in **chapter 4** was the first attempt to evaluate large-scale protein expression patterns in ER-negative breast cancer. This work lays a foundation for understanding protein expression and activation status in these tumors, but the role the discovered proteins play in carcinogenesis has yet to be determined. This study identified proteins and activated pathways that may be important in the process of transformation, potentiation, migration, or metastasis in breast cancer. Future experiments utilizing siRNA and shRNA knockdown technologies will address whether the identified molecules are necessary for mitogenesis, angiogenesis, or invasion. The generation of dominant negative constructs of these molecules will also aid in these functional studies. Xenograft studies using cells that express constitutively active forms of the proteins identified will determine whether activation of these pathways aids or accelerates tumorigenesis. Conversely, small molecule inhibitors will need to be developed and tested in models of ER-negative disease to see if these agents are effective at treating and even preventing breast cancer development.

The identification of the S6 kinase pathway as being overrepresented and activated in ER-negative disease represents another area that merits further exploration. This pathway, implicated previously in cancer, is important for cell survival, growth, translation initiation, and metabolic processes (reviewed in [38]). Interestingly, one of the proteins we found underrepresented in ER-negative disease, TSC2 (a putative tumor suppressor), is a negative regulator of the S6 kinase signaling pathway [39]. Future studies will need to evaluate whether targeting of the S6 kinase pathway, either upstream of S6 kinase with mTOR inhibitors, at the level of S6 kinase itself, or downstream with eIF4 inhibitors, is effective at treating or preventing ER-negative breast cancer. These studies, which will certainly involve both *in vitro* tissue culture and *in vivo* mouse experiments, continue to be a focus of Dr. Brown's laboratory. In fact, this paradigm of "backward screening" will continue to be an important in target validation. This thesis identifies tens, if not hundreds, of potential important molecules in ER-negative breast cancer etiology. All of them will need to be systematically evaluated and validated as effective targets of treatment in ER-negative disease, and will certainly be the focus of experimentation in the years to come. Despite the obvious potential, the promise of novel target discovery has not yet been fulfilled. We contend that this integrative genomic, transcriptional profiling, and functional proteomics approach will more efficiently identify new targets for therapy.

One of the exciting areas of future research involves the translation of scientific discoveries into clinically relevant treatments. The studies described in this thesis

provided some of the preclinical rationale for the development of one such clinical trial. As reported in **chapter 3**, our RNA profiling studies showed that ER-negative tumors have increased expression of various tyrosine kinases, including EGFR, SRC, YES-1, KIT, EPH receptors B4, and ABL compared to ER-positive tumors. Based on these observations, we initiated a phase II preoperative study of dasatinib, which is a potent oral multi-kinase inhibitor that targets ABL, SRC, KIT, PDGFR, and other tyrosine kinases, to treat women with metastatic ER-negative breast cancer. **Table 6.1** lists the targets in ER-negative breast cancer that are inhibited by dasatinib. Dasatinib has already received FDA-approval for the treatment of patients with chronic myeloid leukemia or Ph-positive acute lymphoblastic leukemia [40], but recent data showed that dasatinib selectively inhibits growth of basal-type ER/PR/HER2 negative breast cancer cell lines growing in vitro [41]. Using in vitro assays, 7 (most of which were 'basal like') of 23 well-characterized breast cancer cell lines, were shown to be relatively sensitive to dasatinib (IC50 ≤ 1μM) [41]. These data, validated independently, strongly supported our preliminary data that dasatinib should inhibit a subset of triple negative breast cancers. Thus, a phase II clinical trial was designed and initiated using dasatinib in women with "triple-negative" breast cancer. The primary endpoints in this preoperative trial will be clinical response rate and toxicity in treatment-naïve patients when given as a single agent for two to four weeks. Secondary endpoints are also included and involve measuring proliferation changes, biomarker response, predictive studies, and monitoring response by novel imaging modalities. This clinical trial highlights the true

Table 6.1 – Kinases targeted by Dasatinib		
Gene Name	**Symbol**	**K_D (nm)**
abl-interactor 1	ABI1	0.5
epidermal growth factor receptor	EGFR	100
EPH receptor A4	EPHA4	0.8
EPH receptor B4	EPHB4	0.3
FYN	FYN	0.7
v-kit Hardy-Zuckerman sarcoma viral oncogene homolog	KIT	0.6
lymphocyte-specific protein tyrosine kinase	LCK	0.2
v-yes-1 Yamaguchi sarcoma viral related oncogene	LYN (YES)	0.6
mitogen-activated protein kinase kinase kinase kinase 4	MAP4K4	50
mitogen-activated protein kinase kinase kinase kinase 5	MAP4K5	50
p38 alpha	MAPK14	30
pim-1 oncogene /// pim-1 oncogene	PIM1	300
receptor-interacting serine-threonine kinase 2	RIPK2	30
SRC	SRC	0.2
serine/threonine kinase 38 like	STK38L	200
serine/threonine kinase 38	STK38	200

Table 6.1 – List of kinases that are targeted by Dasatinib. This table lists the kinases with the corresponding Kd values that are both overexpressed in ER-negative breast cancer and are the targets of Dasitinib drug inhibition.

promise of such translational studies, and underscores the true "bench-to-bedside" potential of this work.

Future clinical management of women with breast cancer will employ a multi-faceted approach. Women diagnosed with breast cancer will still be dichotomized based on their estrogen receptor expression status, with patients either being ER-positive or ER-negative. Whereas this is currently the extent of patient classification, future management will include other measures meant to further stratify patients, including the use of DNA, RNA, and proteomic profiling to molecularly profile each individual tumor. Based on this information, the clinician will be able to design a tailored treatment regimen that has been proven efficacious in breast cancer. This may include the continued use of surgical resection and selective estrogen receptor modulators and aromatase inhibitors for women whose tumors are ER-positive and have a Luminal A gene expression profile. These women, then, would have reasonable expectation of a good prognosis and possible cure. Other women who are ER-positive but have a Luminal B gene expression profile may require more aggressive therapy, but with the initiation of this therapy early in the course of disease management, these women will also have a reasonable expectation of a good outcome. Similarly, this profiling of tumors early in the course of disease treatment will benefit women with ER-negative breast cancer. As effective inhibitors for the molecules identified in this thesis are developed, additional treatment options will be available. Depending on the molecules and signature that these ER-negative tumor's have (S6-kinase high, Cell cycle

checkpoint high, etc.), the treatment strategy will change. This ability to target the molecular abnormalities of each tumor individually will certainly lead to better outcomes in patients with breast, including those patients with difficult to treat "triple-negative" breast cancer. **Figure 6.1** outlines this future of breast cancer disease treatment, including the molecularly selected treatment for ER-negative breast cancer. In addition to the therapies currently available for treating breast cancer, future efforts will lead to additional clinical trials evaluating inhibitors against the molecules identified in this thesis, including S6 kinase, PAI1, CHEK1, and MELK, as effective therapies in the treatment of ER-negative disease.

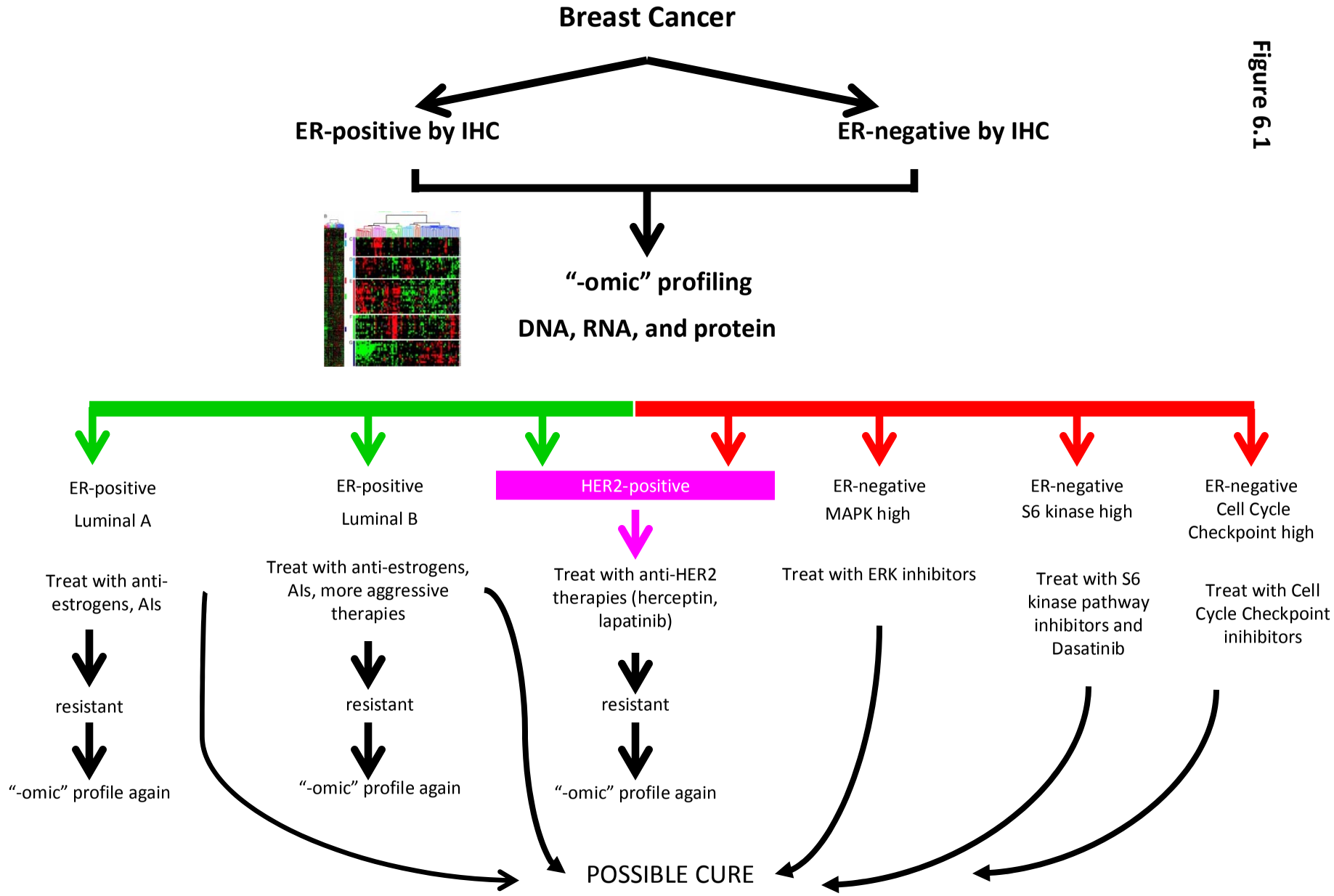

Breast Cancer
ER-positive by IHC
ER-negative by IHC
"-omic" profiling
DNA, RNA, and protein
ER-positive
Luminal A
ER-positive
Luminal B
HER2-positive
ER-negative
MAPK high
ER-negative
S6 kinase high
ER-negative
Cell Cycle
Checkpoint high
Treat with anti-estrogens, AIs
Treat with anti-estrogens, AIs, more aggressive therapies
Treat with anti-HER2 therapies (herceptin, lapatinib)
Treat with ERK inhibitors
Treat with S6 kinase pathway inhibitors and Dasatinib
Treat with Cell Cycle Checkpoint inihibitors
resistant
resistant
resistant
"-omic" profile again
"-omic" profile again
"-omic" profile again
POSSIBLE CURE

Figure 6.1

6.4 Final Remarks

From the early use of hormonal therapy to trastuzumab, targeted therapies have clearly found a place in the treatment of breast cancer. The last decade has seen several new targets and new therapeutic approaches move from the laboratory into clinical trials. As those trials mature, we will determine the impact of these new targeted therapies. Equally exciting is the potential of genomics and proteomics to discover the hidden targets of the current cytotoxic arsenal. In reality, all therapy is targeted -- we just don't know all of the targets yet. The groundwork laid in this thesis, coupled with the continuing biotechnology revolution, holds the promise that effective and tolerable treatments for ER-negative breast cancer will soon be available.

Some argue that this future of breast cancer treatment development relies on the unique combinations of existing drugs. They point to the initial excitement surrounding the use of Herceptin and anthracyclines and point to its initial positive results. Unfortunately, cardiotoxity also emerged as a severe side effect that limited the utility of this drug regimen clinically. While not discounting the role of these potentially useful combinations of pre-existing therapies, more specific targeted therapy may offer more promise. The future of targeted therapy in ER-negative breast cancer is more promising than the mix-and-match approach currently envisioned. In the future it will be possible to treat women with targeted drugs for their particular type of breast cancer using DNA, RNA, and proteomic profiling to select effective drugs at the time of diagnosis. With the development of genome sequencing, gene expression profiling, and

proteomic assessments, coupled with the concomitant decrease in the cost of "-omic" technologies, personalized medicine is no longer a dream. Clinicians will soon have the ability to thoroughly profile a patient's tumor and design a targeted and tailored treatment regimen. With greater understanding of the mechanisms that lead to drug resistance, prognostic and predictive biomarkers, and novel treatment strategies that overcome resistance, clinicians will soon be able to effectively treat ER-negative breast cancer. Breast cancer will no longer be defined in terms of what is NOT known about the tumor (i.e. ER-negative, PR-negative, HER2-negative breast cancer) but by what we DO know (PI3-kinase pathway activated, caveolin high, or PAI-1 high expressing) and will be treated accordingly. With this knowledge and progress, it will soon be possible to effectively treat, and even cure, this now deadly disease.

References:

1. Carey, L.A., et al., *Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.* Jama, 2006. **295**(21): p. 2492-502.
2. Patel, R.R., C.G. Sharma, and V.C. Jordan, *Optimizing the antihormonal treatment and prevention of breast cancer.* Breast Cancer, 2007. **14**(2): p. 113-22.
3. Geyer, C.E., et al., *Lapatinib plus capecitabine for HER2-positive advanced breast cancer.* N Engl J Med, 2006. **355**(26): p. 2733-43.
4. Dowell, J.E. and J.D. Minna, *EGFR mutations and molecularly targeted therapy: a new era in the treatment of lung cancer.* Nat Clin Pract Oncol, 2006. **3**(4): p. 170-1.
5. Hinoda, Y., et al., *Monoclonal antibodies as effective therapeutic agents for solid tumors.* Cancer Sci, 2004. **95**(8): p. 621-5.
6. Miller, K.D. and C.L. Dul, *Breast cancer: the role of angiogenesis and antiangiogenic therapy.* Hematol Oncol Clin North Am, 2004. **18**(5): p. 1071-86, ix.
7. Meyer, K.B., et al., *Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer.* PLoS Biol, 2008. **6**(5): p. e108.
8. Moulder, S.L. and C.L. Arteaga, *A Phase I/II Trial of trastuzumab and gefitinib in patients with Metastatic Breast Cancer that overexpresses HER2/neu (ErbB-2).* Clin Breast Cancer, 2003. **4**(2): p. 142-5.
9. Polychronis, A., et al., *Preoperative gefitinib versus gefitinib and anastrozole in postmenopausal patients with oestrogen-receptor positive and epidermal-growth-factor-receptor-positive primary breast cancer: a double-blind placebo-controlled phase II randomised trial.* Lancet Oncol, 2005. **6**(6): p. 383-91.
10. Tan, A.R., et al., *Evaluation of biologic end points and pharmacokinetics in patients with metastatic breast cancer after treatment with erlotinib, an epidermal growth factor receptor tyrosine kinase inhibitor.* J Clin Oncol, 2004. **22**(15): p. 3080-90.
11. Spector, N.L., et al., *Study of the biologic effects of lapatinib, a reversible inhibitor of ErbB1 and ErbB2 tyrosine kinases, on tumor growth and survival pathways in patients with advanced malignancies.* J Clin Oncol, 2005. **23**(11): p. 2502-12.
12. Teschendorff, A.E., et al., *An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer.* Genome Biol, 2007. **8**(8): p. R157.
13. Amini, R.M., et al., *Mast cells and eosinophils in invasive breast carcinoma.* BMC Cancer, 2007. **7**(1): p. 165.
14. Denardo, D.G. and L.M. Coussens, *Inflammation and breast cancer. Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression.* Breast Cancer Res, 2007. **9**(4): p. 212.

15. Kerbel, R. and J. Folkman, *Clinical translation of angiogenesis inhibitors.* Nat Rev Cancer, 2002. **2**(10): p. 727-39.
16. Climent, J., et al., *Deletion of chromosome 11q predicts response to anthracycline-based chemotherapy in early breast cancer.* Cancer Res, 2007. **67**(2): p. 818-26.
17. Climent, J., et al., *Characterization of breast cancer by array comparative genomic hybridization.* Biochem Cell Biol, 2007. **85**(4): p. 497-508.
18. Johnson, N., et al., *A comparative study of genome-wide SNP, CGH microarray and protein expression analysis to explore genotypic and phenotypic mechanisms of acquired antiestrogen resistance in breast cancer.* Breast Cancer Res Treat, 2008. **111**(1): p. 55-63.
19. Pollack, J.R., et al., *Genome-wide analysis of DNA copy-number changes using cDNA microarrays.* Nat Genet, 1999. **23**(1): p. 41-6.
20. Rodriguez, C., et al., *At least five regions of imbalance on 6q in breast tumors, combining losses and gains.* Genes Chromosomes Cancer, 2000. **27**(1): p. 76-84.
21. Werner, M., et al., *20q13.2 amplification in intraductal hyperplasia adjacent to in situ and invasive ductal carcinoma of the breast.* Virchows Arch, 1999. **435**(5): p. 469-72.
22. Cheng, K.W., et al., *The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers.* Nat Med, 2004. **10**(11): p. 1251-6.
23. Prentice, L.M., et al., *NRG1 gene rearrangements in clinical breast cancer: identification of an adjacent novel amplicon associated with poor prognosis.* Oncogene, 2005. **24**(49): p. 7281-9.
24. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.
25. Wong, K.K., et al., *Allelic imbalance analysis by high-density single-nucleotide polymorphic allele (SNP) array with whole genome amplified DNA.* Nucleic Acids Res, 2004. **32**(9): p. e69.
26. Naylor, T.L., et al., *High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization.* Breast Cancer Res, 2005. **7**(6): p. R1186-98.
27. Rai, R., et al., *BRIT1 regulates early DNA damage response, chromosomal integrity, and cancer.* Cancer Cell, 2006. **10**(2): p. 145-57.
28. Chang, H.Y., et al., *Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds.* PLoS Biol, 2004. **2**(2): p. E7.
29. Chang, J., et al., *Survival of patients with metastatic breast carcinoma: importance of prognostic markers of the primary tumor.* Cancer, 2003. **97**(3): p. 545-53.

30. Chang, J.C., et al., *Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer.* Lancet, 2003. **362**(9381): p. 362-9.
31. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.
32. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
33. Sotiriou, C., et al., *Breast cancer classification and prognosis based on gene expression profiles from a population-based study.* Proc Natl Acad Sci U S A, 2003. **100**(18): p. 10393-8.
34. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.
35. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-6.
36. Chang, J.C., et al., *Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients.* J Clin Oncol, 2005. **23**(6): p. 1169-77.
37. Sotiriou, C., et al., *Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer.* Breast Cancer Res, 2002. **4**(3): p. R3.
38. Inoki, K., M.N. Corradetti, and K.L. Guan, *Dysregulation of the TSC-mTOR pathway in human disease.* Nat Genet, 2005. **37**(1): p. 19-24.
39. Jaeschke, A., et al., *Tuberous sclerosis complex tumor suppressor-mediated S6 kinase inhibition by phosphatidylinositide-3-OH kinase is mTOR independent.* J Cell Biol, 2002. **159**(2): p. 217-24.
40. Talpaz, M., et al., *Dasatinib in imatinib-resistant Philadelphia chromosome-positive leukemias.* N Engl J Med, 2006. **354**(24): p. 2531-41.
41. Finn, R.S., et al., *Dasatinib, an orally active small molecule inhibitor of both the src and abl kinases, selectively inhibits growth of basal-type/"triple-negative" breast cancer cell lines growing in vitro.* Breast Cancer Res Treat, 2007. **105**(3): p. 319-26.